WEST Search History

Hide Items | Restore | Clear | Cancel

DATE: Friday, January 06, 2006

Hide?	Set Name	Query	Hit Count
		DB=PGPB,USPT,EPAB; PLUR=YES; OP=ADJ	
☐	L12	L11 AND L9	84
☐	L11	L10 AND L3	323
☐	L10	L2 AND L1	1146
☐	L9	L8 or L7	29257
☐	L8	(435/7.1 |435/7.23)![CCLS]	13124
☐	L7	(530/350)![CCLS]	18276
☐	L6	(530)![CCLS]	0
☐	L5	(530)![CCLS]	0
☐	L4	2A2A9	0
☐	L3	prostate	38232
☐	L2	tumor$ or tumuor$ or cancer$ or neoplas$	180591
☐	L1	(jakobovits or afar or challita$ or levin or mitchell or hubert).in.	34377

END OF SEARCH HISTORY

OM protein - protein search, using sw model

Run on: December 4, 2005, 09:57:24 ; Search time 187 Seconds
                                              (without alignments)
                                              1184.208 Million cell updates/sec

Title: US-09-771-312-2
Perfect score: 2694
Sequence: 1 MEELVHDLVSALEESSEQAR.........GFPLPKSTSATTTPNAGKSA 504

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5   .

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_21:*
                 1: geneseqp1980s:*
                 2: geneseqp1990s:*
                 3: geneseqp2000s:*
                 4: geneseqp2001s:*
                 5: geneseqp2002s:*
                 6: geneseqp2003as:*
                 7: geneseqp2003bs:*
                 8: geneseqp2004s:*
                 9: geneseqp2005s:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                              SUMMARIES

| | | % Query | | | | |
Result No.	Score	Match	Length	DB	ID	Description
1	2694	100.0	504	4	AAU06524	Aau06524 Prostate
2	2694	100.0	528	4	AAB92632	Aab92632 Human pro
3	2694	100.0	528	5	ABB97288	Abb97288 Novel hum
4	1808	67.1	376	8	ADR99239	Adr99239 Hypotheti
5	825.5	30.6	313	4	ABG08002	Abg08002 Novel hum
6	608.5	22.6	423	4	ABG23408	Abg23408 Novel hum
7	591.5	22.0	446	5	ABB75706	Abb75706 Human pho
8	588.5	21.8	482	5	ABP43772	Abp43772 14 clone

9	533.5	19.8	453	5	ABB97561	Abb97561 Novel hum
10	530.5	19.7	351	4	AAB94662	Aab94662 Human pro
11	530.5	19.7	351	5	ABB97470	Abb97470 Novel hum
12	338	12.5	275	4	AAB92468	Aab92468 Human pro
13	320.5	11.9	223	4	AAM15386	Aam15386 Peptide #
14	320.5	11.9	223	4	ABB34392	Abb34392 Peptide #
15	320.5	11.9	223	4	AAM27874	Aam27874 Peptide #
16	320.5	11.9	223	4	ABB29229	Abb29229 Peptide #
17	320.5	11.9	223	4	AAM67577	Aam67577 Human bon
18	320.5	11.9	223	4	AAM55182	Aam55182 Human bra
19	320.5	11.9	223	4	ABG49223	Abg49223 Human liv
20	320.5	11.9	223	4	AAM03148	Aam03148 Peptide #
21	320.5	11.9	223	5	ABG37168	Abg37168 Human pep
22	320.5	11.9	223	8	ABO59933	Abo59933 Human gen
23	236.5	8.8	123	4	AAM95677	Aam95677 Human rep
24	196	7.3	1038	7	ADC03412	Adc03412 Rice flow
25	196	7.3	1038	7	ABM88777	Abm88777 Rice abio
26	181.5	6.7	564	8	ADY23792	Ady23792 Plant ful
27	167.5	6.2	767	6	ABR53431	Abr53431 Protein s
28	167.5	6.2	767	7	ADK64670	Adk64670 Disease t
29	164.5	6.1	554	3	AAG36165	Aag36165 Arabidops
30	164.5	6.1	652	3	AAG36164	Aag36164 Arabidops
31	164.5	6.1	781	3	AAG36163	Aag36163 Arabidops
32	162.5	6.0	815	5	AAG78388	Aag78388 Human H37
33	162.5	6.0	815	7	AAE38620	Aae38620 Human H37
34	162.5	6.0	815	7	AAE38621	Aae38621 Human H37
35	162.5	6.0	815	8	ADP23184	Adp23184 PRO polyp
36	162.5	6.0	815	9	ADX05546	Adx05546 Cyclin-de
37	162	6.0	381	2	AAY07056	Aay07056 Renal can
38	157.5	5.8	852	7	ADD45318	Add45318 Rat Prote
39	157.5	5.8	852	7	ADE56352	Ade56352 Rat Prote
40	156.5	5.8	573	4	ADM19760	Adm19760 Protein e
41	156	5.8	852	7	AEA79131	Aea79131 Human apo
42	156	5.8	852	9	ADX07612	Adx07612 Cyclin-de
43	156	5.8	929	4	AAM78604	Aam78604 Human pro
44	156	5.8	930	8	ABM82400	Abm82400 Tumour-as
45	156	5.8	930	9	ADX07610	Adx07610 Cyclin-de

ALIGNMENTS

RESULT 1
AAU06524
ID   AAU06524 standard; protein; 504 AA.
XX
AC   AAU06524;
XX
DT   24-OCT-2001 (first entry)
XX
DE   Prostate and testis-related gene 84P2A9 encoded protein.
XX
KW   84P2A9-related protein; prostate; testis; tissue; cancer; leukaemia;
KW   tumour; kidney; brain; bone; skin; ovary; breast; pancreas; colon; lung;
KW   cytostatic; gene therapy; antibody therapy; ribozyme; serum; blood;
KW   single chain monoclonal antibody; urine.
XX

```
  Query Match              100.0%;  Score 2694;  DB 4;  Length 504;
  Best Local Similarity    100.0%;  Pred. No. 3.9e-228;
  Matches  504;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 60


Qy       61 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 120


Qy      121 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 180
```

```
Qy    181  TKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  TKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 240

Qy    241  STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 300

Qy    301  ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPD 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPD 360

Qy    361  SHHHDHWFSPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRR 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  SHHHDHWFSPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRR 420

Qy    421  KAAPLPGPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPK 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  KAAPLPGPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPK 480

Qy    481  GLGLGFPLPKSTSATTTPNAGKSA 504
           ||||||||||||||||||||||||
Db    481  GLGLGFPLPKSTSATTTPNAGKSA 504
```

RESULT 2
AAB92632
ID   AAB92632 standard; protein; 528 AA.
XX
AC   AAB92632;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human protein sequence SEQ ID NO:10938.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX

PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 10938; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 528 AA;

     Query Match   100.0%;   Score 2694; DB 4;   Length 528;
     Best Local Similarity   100.0%;   Pred. No. 4.2e-228;
     Matches   504; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 84

Qy   61 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 144

Qy   121 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 204

Qy   181 TKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   205 TKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 264

Qy   241 STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   265 STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 324

```
Qy     301  ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPD 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325  ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPD 384

Qy     361  SHHHDHWFSPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRR 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     385  SHHHDHWFSPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRR 444

Qy     421  KAAPLPGPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPK 480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     445  KAAPLPGPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPK 504

Qy     481  GLGLGFPLPKSTSATTTPNAGKSA 504
            ||||||||||||||||||||||||
Db     505  GLGLGFPLPKSTSATTTPNAGKSA 528
```

RESULT 3
ABB97288
ID   ABB97288 standard; protein; 528 AA.
XX
AC   ABB97288;
XX
DT   28-JUN-2002 (first entry)
XX
DE   Novel human protein SEQ ID NO: 556.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US026015.
XX
PR   11-SEP-2000; 2000US-00659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   N-PSDB; ABN32474.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis.
XX
PS   Example 2; SEQ ID NO 556; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444

```
  Query Match              100.0%;  Score 2694;  DB 5;  Length 528;
  Best Local Similarity    100.0%;  Pred. No. 4.2e-228;
  Matches  504;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       25 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 84

Qy       61 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       85 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 144

Qy      121 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      145 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 204

Qy      181 TKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      205 TKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 264

Qy      241 STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      265 STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 324

Qy      301 ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPD 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      325 ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPD 384

Qy      361 SHHHDHWFSPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRR 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      385 SHHHDHWFSPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRR 444

Qy      421 KAAPLPGPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPK 480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      445 KAAPLPGPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPK 504

Qy      481 GLGLGFPLPKSTSATTTPNAGKSA 504
            ||||||||||||||||||||||||
Db      505 GLGLGFPLPKSTSATTTPNAGKSA 528
```

RESULT 4
ADR99239
ID   ADR99239 standard; protein; 376 AA.
XX

```
AC    ADR99239;
XX
DT    02-DEC-2004  (first entry)
XX
DE    Hypothetical protein FLJ10252, SEQ ID 245.
XX
KW    Cytostatic; breast cancer; cancer; human; FLJ10252.
XX
OS    Homo sapiens.
XX
PN    WO2004078035-A2.
XX
PD    16-SEP-2004.
XX
PF    27-FEB-2004; 2004WO-US007268.
XX
PR    28-FEB-2003; 2003US-0450655P.
XX
PA    (FARB ) BAYER PHARM CORP.
XX
PI    Eveleigh D,  Bigwood D;
XX
DR    WPI; 2004-653556/63.
DR    N-PSDB; ADR99112.
XX
PT    Diagnosing breast cancer comprises comparing the level of expression of
PT    genes or gene products in a first biological sample taken from a patient
PT    with that in a normal patient sample.
XX
PS    Claim 3; SEQ ID NO 245; 53pp; English.
XX
CC    The present invention relates to a method (M1) for diagnosing breast
CC    cancer in a patient. The method comprises comparing the level of
CC    expression of one or more genes or gene products in a biological sample
CC    from the patient with that in a normal patient sample, where a difference
CC    in the gene expression in the first sample compared to that in the second
CC    sample is a diagnostic of the disease. Also claimed are: method (M2) for
CC    distinguishing between normal and disease tissues; method (M3) for
CC    monitoring the response of a breast cancer patient to treatment with an
CC    anti-cancer agent; method (M4) for identifying a compound for treating
CC    breast cancer; and an array for distinguishing between normal and disease
CC    tissues comprising two or more probes corresponding to genes selected
CC    from ADR98995-ADR99121 or comprising two or more polypeptides selected
CC    from ADR99122-ADR99248. In M1 and M2 the genes are selected from ADR98995
CC    -ADR99121 and the gene products are polypeptides selected from ADR99122-
CC    ADR99248. M1 is useful for diagnosing breast cancer. M2 and the array are
CC    useful for distinguishing between normal and disease tissue. M3 is useful
CC    for monitoring the response of a breast cancer patient to treatment with
CC    an anti-cancer agent. M4 is useful for identifying a compound for
CC    treating breast cancer. Note: The sequence data for this patent did not
CC    form part of the printed specification, but was obtained in electronic
CC    format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ    Sequence 376 AA;

   Query Match            67.1%;  Score 1808;  DB 8;  Length 376;
   Best Local Similarity  99.7%;  Pred. No. 3.1e-150;
```

```
     Matches   341;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       25 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 84

Qy       61 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       85 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 144

Qy      121 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      145 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 204

Qy      181 TKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 240
            |||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db      205 TKNKVKKRKLKIIRQGPKIQNEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 264

Qy      241 STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      265 STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 324

Qy      301 ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSM 342
            ||||||||||||||||||||||||||||||||||||||||||
Db      325 ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSM 366
```

RESULT 5
ABG08002

CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 313 AA;

   Query Match   30.6%; Score 825.5; DB 4;   Length 313;
   Best Local Similarity   55.7%; Pred. No. 9.2e-64;
   Matches 181; Conservative   4; Mismatches   25; Indels 115; Gaps   5;

Qy   1 MEELVHDLVSALE-ESSEQARGGFAETGD-HSRSISCPLKRQARKRRGRKRRSYNVHHPW 58
           ||||||||||||| | : |||||| || | ||||| ||||||||||| |
Db   37 MEELVHDLVSALERELQSKPRGGFAEPGDPFSEVYPCPLKRPARKRRGRKRRFVXCASP- 95

Qy   59 ETGHCL---SEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGK 115
           | : |||||||
Db   96 -VGGLVTAXSEGSDSS--- 110

Qy   116 RPLWHESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAY 175

Db   111 --- 110

Qy   176 QYQEFTKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESD 235
                    | : | |||||||||||||||||||||||||||||||||||||||||||||||
Db   111 ----FRRTKSKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESD 166

Qy   236 SSSLSSTDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPD 295
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   167 SSSLSSTDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPD 226

Qy   296 PVFESILTGSFPLMSHPSRRGFQAR 320

```
         ||||||||||||||||||||||  :
Db   227 PVFESILTGSFPLMSHPSRRGFPTK 251
```

```
                        GenCore version 5.1.6
                Copyright (c) 1993 - 2005  Compugen Ltd.


OM protein - nucleic search, using frame_plus_p2n model

Run on:           December 11, 2005, 17:36:59 ; Search time 7575 Seconds
                                     (without alignments)
                                     3782.059 Million cell updates/sec


Title:            US-09-771-312-2
Perfect score:    2694
Sequence:         1 MEELVHDLVSALEESSEQAR..........GFPLPKSTSATTTPNAGKSA 504


Scoring table:    BLOSUM62
                  Xgapop 10.0 , Xgapext   0.5
                  Ygapop 10.0 , Ygapext   0.5
                  Fgapop  6.0 , Fgapext   7.0
                  Delop   6.0 , Delext    7.0


Searched:         5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters:     11766282


Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09771312/runat_01122005_145311_15042/app_query.fasta_
1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09771312_@CGN_1_1_4939_@runat_01122005_145311_15042 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :        GenEmbl:*
                  1:  gb_ba:*
                  2:  gb_in:*
                  3:  gb_env:*
                  4:  gb_om:*
                  5:  gb_ov:*
                  6:  gb_pat:*
                  7:  gb_ph:*
                  8:  gb_pr:*
                  9:  gb_ro:*
                  10:  gb_sts:*
```

```
     11:   gb_sy:*
     12:   gb_un:*
     13:   gb_vi:*
     14:   gb_htg:*
     15:   gb_pl:*


Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                                  SUMMARIES
                         %
Result                 Query
     No.    Score   Match Length  DB    ID                    Description
       -------------------------------------------------------------------------
        1    2694   100.0   2338   6    BD155908              BD155908 Primer fo
        2    2694   100.0   2338   6    AX876032              AX876032 Sequence
        3    2694   100.0   2338   8    AK001114              AK001114 Homo sapi
        4    2694   100.0   2344   6    AX405697              AX405697 Sequence
        5    2694   100.0   2345   6    AX206855              AX206855 Sequence
        6    2318    86.0   4537   9    BC054810              BC054810 Mus muscu
        7  2014.5    74.8   5187   5    AJ851518              AJ851518 Gallus ga
        8    1813    67.3   1026   6    CQ720787              CQ720787 Sequence
        9    1813    67.3   4022   6    BD183390              BD183390 Novel gen
       10    1808    67.1   3189   8    BC042193              BC042193 Homo sapi
       11    1807    67.1   3250   8    BC063474              BC063474 Homo sapi
       12  1587.5    58.9   1350   5    BC097745              BC097745 Xenopus 1
       13  1513.5    56.2   1392   9    BC079232              BC079232 Rattus no
  c    14    1229    45.6 135060   8    AL354659              AL354659 Human DNA
  c    15    1229    45.6 142908  14    AL513172               AL513172 Homo sapi
       16  1072.5    39.8 180315   9    AC107843              AC107843 Mus muscu
  c    17  1072.5    39.8 260404   9    AC110033              AC110033 Mus muscu
       18    1054    39.1    817   6    BD146304              BD146304 Primer fo
       19    1054    39.1    817   6    AX866242              AX866242 Sequence
       20    1038    38.5 254644  14    AC136836               AC136836 Rattus no
       21    1038    38.5 256511  14    AC135040               AC135040 Rattus no
  c    22    1038    38.5 262721  14    AC106265               AC106265 Rattus no
       23    1008    37.4    759   8    BC027719              BC027719 Homo sapi
       24   627.5    23.3   3947   9    BC048169              BC048169 Mus muscu
       25   627.5    23.3   3947   9    BC058256              BC058256 Mus muscu
       26   627.5    23.3   4311   9    AK129299              AK129299 Mus muscu
       27   627.5    23.3   4314   9    BC050782              BC050782 Mus muscu
       28     608    22.6   1021   8    AK024701              AK024701 Homo sapi
  c    29   566.5    21.0 148801   5    BX004824              BX004824 Zebrafish
  c    30   566.5    21.0 151096  14    BX005303               BX005303 Danio rer
       31   563.5    20.9   2434   6    AX405970              AX405970 Sequence
       32   563.5    20.9   2463   8    AK000696              AK000696 Homo sapi
       33     544    20.2    407   6    CQ735676              CQ735676 Sequence
       34     534    19.8    469   6    BD108636              BD108636 EST and e
       35     534    19.8    469   6    AR413083              AR413083 Sequence
       36     534    19.8    469   6    AX969917              AX969917 Sequence
       37   530.5    19.7   1474   6    AX405879              AX405879 Sequence
       38   530.5    19.7   1485   8    BC058032              BC058032 Homo sapi
       39   530.5    19.7   2112   6    BD158526              BD158526 Primer fo
       40   530.5    19.7   2112   6    AX880680              AX880680 Sequence
       41   530.5    19.7   2112   8    AK023523              AK023523 Homo sapi
```

```
42    530.5   19.7    6256   8   AB032978              AB032978 Homo sapi
43    459.5   17.1     526   5   CT025336              CT025336 Xenopus t
44      422   15.7    1490   8   BC038835              BC038835 Homo sapi
45      420   15.6     445   6   CQ431223              CQ431223 Sequence
```

ALIGNMENTS

RESULT 1
BD155908
LOCUS   BD155908   2338 bp   DNA   linear   PAT 17-JAN-
2003
DEFINITION   Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD155908
VERSION   BD155908.1 GI:27861666
KEYWORDS   JP 2002191363-A/10751.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Hominidae; Homo.
REFERENCE   1 (bases 1 to 2338)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K.,
Yamamoto,J.,
                Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE   Primer for synthesizing full-length cDNA and use thereof
  JOURNAL   Patent: JP 2002191363-A 10751 09-JUL-2002;
                HELIX RESEARCH INSTITUTE
COMMENT   OS   Homo sapiens (human)
                PN   JP 2002191363-A/10751
                PD   09-JUL-2002
                PF   28-JUL-2000 JP 2000280990
                PI   TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
                PI   SAITO,
                PI   JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI
WAKAMATSU,
                PI   KEIICHI NAGAI,TETSUJI OTSUKI
                PC
                C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
                 10,
                PC   C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
                 Primer for synthesizing full-length cDNA and use thereof FH   Key
                      Location/Qualifiers
                FT   CDS   (99)..(1682).
FEATURES   Location/Qualifiers
     source   1..2338
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:   5.58e-138   Length: 2338
Score: 2694.00   Matches: 504
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
```

Query Match:              100.00%        Indels:         0
DB:                       6              Gaps:           0

US-09-771-312-2 (1-504) x BD155908 (1-2338)

```
Qy        1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      171 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTGGAAGAGAGCTCAGAGCAAGCTCGA 230

Qy       21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      231 GGTGGATTTGCTGAAACAGGAGACCATTCTCGAAGTATATCTTGCCCTCTGAAACGCCAG 290

Qy       41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      291 GCAAGGAAAAGGAGAGGGAGAAAACGGAGGTCGTATAATGTGCATCACCCGTGGGAGACT 350

Qy       61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      351 GGTCACTGCTTAAGTGAAGGCTCTGATTCTAGTTTAGAAGAACCAAGCAAGGACTATAGA 410

Qy       81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      411 GAGAATCACAATAATAATAAAAAAGATCACAGTGACTCTGATGACCAAATGTTAGTAGCA 470

Qy      101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      471 AAGCGCAGGCCGTCATCAAACTTAAATAATAATGTTCGAGGGAAAAGACCTCTATGGCAT 530

Qy      121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      531 GAGTCTGATTTTGCTGTGGACAATGTTGGGAATAGAACTCTGCGCAGGAGGAGAAAGGTA 590

Qy      141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      591 AAACGCATGGCAGTAGATCTCCCACAGGACATCTCTAACAAACGGACAATGACCCAGCCA 650

Qy      161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      651 CCTGAGGGTTGTAGAGATCAGGACATGGACAGTGATAGAGCCTACCAGTATCAAGAATTT 710

Qy      181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      711 ACCAAGAACAAAGTCAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCAAAAATCCAA 770

Qy      201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      771 GATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGACAAAATGGAA 830

Qy      221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      831 TGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGCAGTCTCAGC 890

Qy      241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      891 AGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGATGAACAGAGT 950
```

```
Qy       261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       951 GACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTTGTGCCCTGG
1010

Qy       281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1011 TGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTCTTTGAAAGT
1070

Qy       301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1071 ATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTCCAAGCTAGA
1130

Qy       321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1131 CTCAGTCGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAGGGACTCCAACT
1190

Qy       341 SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp 360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1191 TCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATTTTTCCCCGGAT
1250

Qy       361 SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln 380
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1251 TCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATGACCAGCATCAG
1310

Qy       381 LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla 400
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1311 CTTCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTCTGTGAGAACAGCC
1370

Qy       401 SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg 420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1371 AGCAGGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCAAACGGAGAAGA
1430

Qy       421 LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro 440
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1431 AAAGCTGCACCTTTGCCTGGACCTACTACTGCAGGATTTGTAGGTGAAAATGCCCAGCCA
1490

Qy       441 IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer 460
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1491 ATCCTAGAAAATAATATTGGAAACCGAATGCTTCAGAATATGGGCTGGACGCCTGGGTCA
1550

Qy       461 GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys 480
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1551 GGCCTTGGACGAGATGGCAAGGGGATCTCTGAGCCAATTCAAGCCATGCAGAGGCCAAAG
1610
```

```
Qy       481  GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1611  GGATTAGGACTTGGATTTCCTCTACCAAAAAGTACTTCCGCAACTACTACCCCCAATGCA
1670


Qy       501  GlyLysSerAla 504
              ||||||||||||
Db      1671  GGAAAATCCGCC 1682
```

RESULT 2
AX876032

```
LOCUS       AX876032               2338 bp    DNA        linear   PAT 17-DEC-
2003
DEFINITION  Sequence 10937 from Patent EP1074617.
ACCESSION   AX876032
VERSION     AX876032.1  GI:40030768
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K.,
Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primers for synthesising full-length cDNA and their use
  JOURNAL   Patent: EP 1074617-A 10937 07-FEB-2001;
            Research Association for Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1. .2338
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             99. .1685
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAE89098.1"
                     /db_xref="GI:40030769"
```

/translation="MFGAAGRQPIGAPAAGNSWHFSRTMEELVHDLVSALEESSEQAR

GGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWETGHCLSEGSDSSLEEPSKD

YRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWHESDFAVDNVGNRTLRR

RRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEFTKNKVKKRKLKIIR

QGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLSSTDAGLFTNDEG

RQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFESILTGSFPLMS

HPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPDSHHHDHWF

SPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRRKAAPLP

GPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPKGLGL
                    GFPLPKSTSATTTPNAGKSA"
ORIGIN

Alignment Scores:
Pred. No.:   5.58e-138   Length: 2338
Score: 2694.00   Matches: 504
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 6   Gaps: 0

US-09-771-312-2 (1-504) x AX876032 (1-2338)

Qy   1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   171 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTGGAAGAGAGCTCAGAGCAAGCTCGA 230

Qy   21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   231 GGTGGATTTGCTGAAACAGGAGACCATTCTCGAAGTATATCTTGCCCTCTGAAACGCCAG 290

Qy   41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   291 GCAAGGAAAAGGAGAGGGAGAAAACGGAGGTCGTATAATGTGCATCACCCGTGGGAGACT 350

Qy   61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   351 GGTCACTGCTTAAGTGAAGGCTCTGATTCTAGTTTAGAAGAACCAAGCAAGGACTATAGA 410

Qy   81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   411 GAGAATCACAATAATAATAAAAAAGATCACAGTGACTCTGATGACCAAATGTTAGTAGCA 470

Qy   101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   471 AAGCGCAGGCCGTCATCAAACTTAAATAATAATGTTCGAGGGAAAAGACCTCTATGGCAT 530

Qy   121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   531 GAGTCTGATTTTGCTGTGGACAATGTTGGGAATAGAACTCTGCGCAGGAGGAGAAAGGTA 590

Qy   141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   591 AAACGCATGGCAGTAGATCTCCCACAGGACATCTCTAACAAACGGACAATGACCCAGCCA 650

Qy   161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   651 CCTGAGGGTTGTAGAGATCAGGACATGGACAGTGATAGAGCCTACCAGTATCAAGAATTT 710

Qy   181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   711 ACCAAGAACAAAGTCAAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCAAAAATCCAA 770

Qy   201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db          771 GATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGACAAAATGGAA 830

Qy          221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          831 TGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGCAGTCTCAGC 890

Qy          241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          891 AGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGATGAACAGAGT 950

Qy          261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          951 GACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTTGTGCCCTGG
           1010

Qy          281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1011 TGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTCTTTGAAAGT
           1070

Qy          301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1071 ATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTCCAAGCTAGA
           1130

Qy          321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1131 CTCAGTCGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAGGGACTCCAACT
           1190

Qy          341 SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1191 TCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATTTTTCCCCGGAT
           1250

Qy          361 SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln 380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1251 TCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATGACCAGCATCAG
           1310

Qy          381 LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla 400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1311 CTTCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTCTGTGAGAACAGCC
           1370

Qy          401 SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1371 AGCAGGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCAAACGGAGAAGA
           1430

Qy          421 LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro 440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1431 AAAGCTGCACCTTTGCCTGGACCTACTACTGCAGGATTTGTAGGTGAAAATGCCCAGCCA
           1490

Qy          441 IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer 460
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1491 ATCCTAGAAAATAATATTGGAAACCGAATGCTTCAGAATATGGGCTGGACGCCTGGGTCA
1550


Qy       461 GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1551 GGCCTTGGACGAGATGGCAAGGGGATCTCTGAGCCAATTCAAGCCATGCAGAGGCCAAAG
1610


Qy       481 GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla 500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1611 GGATTAGGACTTGGATTTCCTCTACCAAAAAGTACTTCCGCAACTACTACCCCCAATGCA
1670


Qy       501 GlyLysSerAla 504
            ||||||||||||
Db      1671 GGAAAATCCGCC 1682
```

RESULT 3
AK001114

```
LOCUS           AK001114                2338 bp    mRNA    linear    PRI 30-JAN-
2004
DEFINITION      Homo sapiens cDNA FLJ10252 fis, clone HEMBB1000807.
ACCESSION       AK001114
VERSION         AK001114.1  GI:7022173
KEYWORDS        oligo capping; fis (full insert sequence).
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Hominidae; Homo.
REFERENCE       1
  AUTHORS       Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
                Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
                Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
                Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
                Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
                Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
                Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
                Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K.,
                Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,
                Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
                Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S.,
Watanabe,S.,
                Yosida,M., Hotuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
                Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
                Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H.,
Oshima,A.,
                Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,
                Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,
                Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,
                Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,
                Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
                Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
                Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,
                Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T.,
```

Kobatake,N.,
            Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T.,
            Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K.,
            Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T.,
            Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,
            Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
            Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R.,
            Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
  TITLE   Complete sequencing and characterization of 21,243 full-length
            human cDNAs
  JOURNAL   Nat. Genet. 36 (1), 40-45 (2004)
   PUBMED   14702039
REFERENCE   2
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 2338)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-
3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone
selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES   Location/Qualifiers
     source   1. .2338
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HEMBB1000807"
                     /tissue_type="whole embryo, mainly body"
                     /clone_lib="HEMBB1"
                     /dev_stage="embryo, 10 weeks"
                     /note="cloning vector: pME18SFL3"
     CDS   99. .1685
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA91509.1"
                     /db_xref="GI:7022174"

/translation="MFGAAGRQPIGAPAAGNSWHFSRTMEELVHDLVSALEESSEQAR

GGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWETGHCLSEGSDSSLEEPSKD

YRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWHESDFAVDNVGNRTLRR

RRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEFTKNKVKKRKLKIIR

QGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLSSTDAGLFTNDEG

RQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFESILTGSFPLMS

HPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPDSHHHDHWF

SPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRRKAAPLP

GPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPKGLGL
                        GFPLPKSTSATTTPNAGKSA"
ORIGIN

Alignment Scores:
Pred. No.:   5.58e-138   Length: 2338
Score: 2694.00   Matches: 504
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 8   Gaps: 0

US-09-771-312-2 (1-504) x AK001114 (1-2338)

Qy   1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   171 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTGGAAGAGAGCTCAGAGCAAGCTCGA 230

Qy   21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   231 GGTGGATTTGCTGAAACAGGAGACCATTCTCGAAGTATATCTTGCCCTCTGAAACGCCAG 290

Qy   41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   291 GCAAGGAAAAGGAGAGGGAGAAAACGGAGGTCGTATAATGTGCATCACCCGTGGGAGACT 350

Qy   61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg 80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   351 GGTCACTGCTTAAGTGAAGGCTCTGATTCTAGTTTAGAAGAACCAAGCAAGGACTATAGA 410

Qy   81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   411 GAGAATCACAATAATAATAAAAAAGATCACAGTGACTCTGATGACCAAATGTTAGTAGCA 470

Qy   101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   471 AAGCGCAGGCCGTCATCAAACTTAAATAATAATGTTCGAGGGAAAAGACCTCTATGGCAT 530

Qy   121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   531 GAGTCTGATTTTGCTGTGGACAATGTTGGGAATAGAACTCTGCGCAGGAGGAGAAAGGTA 590

Qy   141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160

```
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       591 AAACGCATGGCAGTAGATCTCCCACAGGACATCTCTAACAAACGGACAATGACCCAGCCA 650

Qy       161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       651 CCTGAGGGTTGTAGAGATCAGGACATGGACAGTGATAGAGCCTACCAGTATCAAGAATTT 710

Qy       181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       711 ACCAAGAACAAAGTCAAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCAAAAATCCAA 770

Qy       201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       771 GATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGACAAAATGGAA 830

Qy       221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       831 TGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGCAGTCTCAGC 890

Qy       241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       891 AGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGATGAACAGAGT 950

Qy       261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       951 GACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTTGTGCCCTGG 
        1010

Qy       281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1011 TGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTCTTTGAAAGT 
        1070

Qy       301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1071 ATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTCCAAGCTAGA 
        1130

Qy       321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1131 CTCAGTCGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAGGGACTCCAACT 
        1190

Qy       341 SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp 360
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1191 TCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATTTTTCCCCGGAT 
        1250

Qy       361 SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln 380
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1251 TCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATGACCAGCATCAG 
        1310

Qy       381 LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla 400
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1311 CTTCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTCTGTGAGAACAGCC 
```

1370

```
Qy        401 SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1371 AGCAGGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCAAACGGAGAAGA
1430

Qy        421 LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1431 AAAGCTGCACCTTTGCCTGGACCTACTACTGCAGGATTTGTAGGTGAAAATGCCCAGCCA
1490

Qy        441 IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1491 ATCCTAGAAAATAATATTGGAAACCGAATGCTTCAGAATATGGGCTGGACGCCTGGGTCA
1550

Qy        461 GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1551 GGCCTTGGACGAGATGGCAAGGGGATCTCTGAGCCAATTCAAGCCATGCAGAGGCCAAAG
1610

Qy        481 GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1611 GGATTAGGACTTGGATTTCCTCTACCAAAAAGTACTTCCGCAACTACTACCCCAATGCA
1670

Qy        501 GlyLysSerAla 504
              ||||||||||||
Db       1671 GGAAAATCCGCC 1682
```

RESULT 4
AX405697
LOCUS   AX405697   2344 bp   DNA   linear   PAT 14-JUN-
2002
DEFINITION   Sequence 112 from Patent WO0222660.
ACCESSION   AX405697
VERSION   AX405697.1 GI:21438833
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tang,Y.T., Liu,C., Zhou,P., Asundi,V., Zhang,J., Zhao,Q.A.,
Ren,F.,
                Xue,A.J., Yang,Y., Wehrman,T. and Drmanac,R.T.
  TITLE   Novel nucleic acids and polypeptides
  JOURNAL   Patent: WO 0222660-A 112 21-MAR-2002;
                HYSEQ, INC. (US)
FEATURES   Location/Qualifiers
     source   1. .2344
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"

```
     CDS               98. .1684
                       /note="unnamed protein product"
                       /codon_start=1
                       /protein_id="CAD34804.1"
                       /db_xref="GI:21438834"

/translation="MFGAAGRQPIGAPAAGNSWHFSRTMEELVHDLVSALEESSEQAR

GGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWETGHCLSEGSDSSLEEPSKD

YRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWHESDFAVDNVGNRTLRR

RRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEFTKNKVKKRKLKIIR

QGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLSSTDAGLFTNDEG

RQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFESILTGSFPLMS

HPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPDSHHHDHWF

SPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRRKAAPLP

GPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPKGLGL
                       GFPLPKSTSATTTPNAGKSA"
ORIGIN


Alignment Scores:
Pred. No.:               5.59e-138      Length:          2344
Score:                   2694.00        Matches:         504
Percent Similarity:      100.00%        Conservative:    0
Best Local Similarity:   100.00%        Mismatches:      0
Query Match:             100.00%        Indels:          0
DB:                      6              Gaps:            0

US-09-771-312-2 (1-504) x AX405697 (1-2344)


Qy        1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTGGAAGAGAGCTCAGAGCAAGCTCGA 229

Qy       21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      230 GGTGGATTTGCTGAAACAGGAGACCATTCTCGAAGTATATCTTGCCCTCTGAAACGCCAG 289

Qy       41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      290 GCAAGGAAAAGGAGAGGGAGAAAACGGAGGTCGTATAATGTGCATCACCCGTGGGAGACT 349

Qy       61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg 80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      350 GGTCACTGCTTAAGTGAAGGCTCTGATTCTAGTTTAGAAGAACCAAGCAAGGACTATAGA 409

Qy       81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      410 GAGAATCACAATAATAATAAAAAAGATCACAGTGACTCTGATGACCAAATGTTAGTAGCA 469
```

```
Qy      101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      470 AAGCGCAGGCCGTCATCAAACTTAAATAATAATGTTCGAGGGAAAAGACCTCTATGGCAT 529

Qy      121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      530 GAGTCTGATTTTGCTGTGGACAATGTTGGGAATAGAACTCTGCGCAGGAGGAGAAAGGTA 589

Qy      141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      590 AAACGCATGGCAGTAGATCTCCCACAGGACATCTCTAACAAACGGACAATGACCCAGCCA 649

Qy      161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
            |||||||||||||||||||||||||||||||||||||||||||||||.||||||||||||
Db      650 CCTGAGGGTTGTAGAGATCAGGACATGGACAGTGATAGAGCCTACCAGTATCAAGAATTT 709

Qy      181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      710 ACCAAGAACAAAGTCAAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCAAAAATCCAA 769

Qy      201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      770 GATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGACAAAATGGAA 829

Qy      221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      830 TGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGCAGTCTCAGC 889

Qy      241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      890 AGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGATGAACAGAGT 949

Qy      261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      950 GACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTTGTGCCCTGG
       1009

Qy      281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1010 TGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTCTTTGAAAGT
       1069

Qy      301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1070 ATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTCCAAGCTAGA
       1129

Qy      321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1130 CTCAGTCGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAGGGACTCCAACT
       1189

Qy      341 SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1190 TCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATTTTTCCCCGGAT
       1249
```

```
Qy        361  SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln 380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1250  TCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATGACCAGCATCAG
         1309

Qy        381  LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla 400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1310  CTTCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTCTGTGAGAACAGCC
         1369

Qy        401  SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1370  AGCAGGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCAAACGGAGAAGA
         1429

Qy        421  LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro 440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1430  AAAGCTGCACCTTTGCCTGGACCTACTACTGCAGGATTTGTAGGTGAAAATGCCCAGCCA
         1489

Qy        441  IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer 460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1490  ATCCTAGAAAATAATATTGGAAACCGAATGCTTCAGAATATGGGCTGGACGCCTGGGTCA
         1549

Qy        461  GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1550  GGCCTTGGACGAGATGGCAAGGGGATCTCTGAGCCAATTCAAGCCATGCAGAGGCCAAAG
         1609

Qy        481  GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla 500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1610  GGATTAGGACTTGGATTTCCTCTACCAAAAAGTACTTCCGCAACTACTACCCCCAATGCA
         1669

Qy        501  GlyLysSerAla 504
               ||||||||||||
Db       1670  GGAAAATCCGCC 1681
```

RESULT 5
AX206855

FEATURES   Location/Qualifiers
    source   1. .2345
                    /organism="Homo sapiens"
                    /mol_type="unassigned DNA"
                    /db_xref="taxon:9606"
    CDS   163. .1677
                    /note="unnamed protein product"
                    /codon_start=1
                    /protein_id="CAC60223.1"
                    /db_xref="GI:15394682"

/translation="MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKR

RGRKRRSYNVHHPWETGHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKR

RPSSNLNNNVRGKRPLWHESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQP

PEGCRDQDMDSDRAYQYQEFTKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDK

MECEEQKVSDELMSESDSSSLSSTDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITG

VVPWWEKEDPTELDKNVPDPVFESILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKK

SGGTPTSMVPIPGPVGNKRMVHFSPDSHHHDHWFSPGARTEHDQHQLLRDNRAERGHK

KNCSVRTASRQTSMHLGSLCTGDIKRRRKAAPLPGPTTAGFVGENAQPILENNIGNRM
              LQNMGWTPGSGLGRDGKGISEPIQAMQRPKGLGLGFPLPKSTSATTTPNAGKSA"
ORIGIN

Alignment Scores:
Pred. No.:   5.6e-138   Length: 2345
Score: 2694.00   Matches: 504
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 6   Gaps: 0

US-09-771-312-2 (1-504) x AX206855 (1-2345)

Qy   1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   163 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTGGAAGAGAGCTCAGAGCAAGCTCGA 222

Qy   21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   223 GGTGGATTTGCTGAAACAGGAGACCATTCTCGAAGTATATCTTGCCCTCTGAAACGCCAG 282

Qy   41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   283 GCAAGGAAAAGGAGAGGGAGAAAACGGAGGTCGTATAATGTGCATCACCCGTGGGAGACT 342

Qy   61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg 80

```
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     343 GGTCACTGCTTAAGTGAAGGCTCTGATTCTAGTTTAGAAGAACCAAGCAAGGACTATAGA 402

Qy      81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     403 GAGAATCACAATAATAATAAAAAAGATCACAGTGACTCTGATGACCAAATGTTAGTAGCA 462

Qy     101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     463 AAGCGCAGGCCGTCATCAAACTTAAATAATAATGTTCGAGGGAAAAGACCTCTATGGCAT 522

Qy     121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     523 GAGTCTGATTTTGCTGTGGACAATGTTGGGAATAGAACTCTGCGCAGGAGGAGAAAGGTA 582

Qy     141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     583 AAACGCATGGCAGTAGATCTCCCACAGGACATCTCTAACAAACGGACAATGACCCAGCCA 642

Qy     161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     643 CCTGAGGGTTGTAGAGATCAGGACATGGACAGTGATAGAGCCTACCAGTATCAAGAATTT 702

Qy     181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     703 ACCAAGAACAAAGTCAAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCAAAAATCCAA 762

Qy     201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     763 GATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGACAAAATGGAA 822

Qy     221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     823 TGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGCAGTCTCAGC 882

Qy     241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     883 AGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGATGAACAGAGT 942

Qy     261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     943 GACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTTGTGCCCTGG
Db    1002

Qy     281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1003 TGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTCTTTGAAAGT
Db    1062

Qy     301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1063 ATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTCCAAGCTAGA
Db    1122

Qy     321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      1123  CTCAGTCGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAGGGACTCCAACT
1182

Qy       341  SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp  360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1183  TCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATTTTTCCCCGGAT
1242

Qy       361  SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln  380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1243  TCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATGACCAGCATCAG
1302

Qy       381  LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla  400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1303  CTTCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTCTGTGAGAACAGCC
1362

Qy       401  SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1363  AGCAGGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCAAACGGAGAAGA
1422

Qy       421  LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro  440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1423  AAAGCTGCACCTTTGCCTGGACCTACTACTGCAGGATTTGTAGGTGAAAATGCCCAGCCA
1482

Qy       441  IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer  460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1483  ATCCTAGAAAATAATATTGGAAACCGAATGCTTCAGAATATGGGCTGGACGCCTGGGTCA
1542

Qy       461  GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1543  GGCCTTGGACGAGATGGCAAGGGGATCTCTGAGCCAATTCAAGCCATGCAGAGGCCAAAG
1602

Qy       481  GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla  500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1603  GGATTAGGACTTGGATTTCCTCTACCAAAAAGTACTTCCGCAACTACTACCCCCAATGCA
1662

Qy       501  GlyLysSerAla  504
              ||||||||||||
Db      1663  GGAAAATCCGCC  1674
```

RESULT 6
BC054810
LOCUS   BC054810   4537 bp   mRNA   linear   ROD 30-JUN-
2004
DEFINITION Mus musculus G patch domain containing 2, mRNA (cDNA clone
            MGC:65681 IMAGE:6839419), complete cds.
ACCESSION   BC054810
VERSION   BC054810.1 GI:32452009
KEYWORDS   MGC.

SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4537)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M.,
Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  PUBMED   12477932
REFERENCE   2 (bases 1 to 4537)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (01-JUL-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, University of Iowa
            cDNA Library Preparation: M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
            Thomas L. Casavant.
            Web site: http://genome.uiowa.edu
            Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
            Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,
            Fishler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
            Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J.,
            Casavant,T., Soares,M.B.

            Clone distribution: MGC clone distribution information can be
found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: Plate: Row: Column: 0.
FEATURES   Location/Qualifiers

```
source          1. .4537
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6"
                /db_xref="taxon:10090"
                /clone="MGC:65681 IMAGE:6839419"
                /tissue_type="Brain"
                /clone_lib="NIH_BMAP_GH0"
                /lab_host="DH10B"
                /note="Vector: pYX-ASC"
gene            1. .4537
                /gene="Gpatc2"
                /db_xref="GeneID:67769"
                /db_xref="MGI:1915019"
CDS             86. .1669
                /gene="Gpatc2"
                /codon_start=1
                /product="G patch domain containing 2"
                /protein_id="AAH54810.1"
                /db_xref="GI:32452010"
                /db_xref="GeneID:67769"
                /db_xref="MGI:1915019"
```

/translation="MFGADGRPAIGTAAGKSWHFSRTMEELVHDLVSALEESSEQARG

GFAETGEHSRNLSCPLKRQARKRRGRKRRSYNVHHPWETGHCLSEGSDSSLEEPSKDY

REKHSNNKKDRSDSDDQMLVAKRRPSSNLSSSVRGKRLLWHESDFAVDSLGNRTLRRR

RKVKRMAVDLPQDVSSKRTMTQLPEGCRDQDMDNDRASQYPEFTRKKVKKRKLKGIRP

GPKTQEEGGVLESEERSQPNKDRMEYEEQKASDELRSESDTSSLSSTDAGLFTNDEGR

QGDDEQSDWFYEKESGGACGIAGVVPWWEKDEPAELDTNLPDPVFESILSGSFPLMSH

PGRGGFQARLSRLHGTPSKNIKKSSGAPPSMLPAPGPGSNKRMVHFSPDAHRHDHWFS

PGARTEHGQHQLLRDNRAERGHKKSCSLKTASRQTSMHLGSLCTGDIKRRRKAAPLPG

PTAAGIVGENAQPILESNIGNRMLQSMGWTPGSGLGRDGRGIAEPVQAVQRPKGLGLG
                        FPLPKSSPTSPAPTSGNPA"
ORIGIN

Alignment Scores:
Pred. No.:   3.54e-117   Length: 4537
Score: 2318.00   Matches: 429
Percent Similarity: 92.06%   Conservative: 35
Best Local Similarity: 85.12%   Mismatches: 40
Query Match: 86.04%   Indels: 0
DB: 9   Gaps: 0

US-09-771-312-2 (1-504) x BC054810 (1-4537)

Qy   1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   155 ATGGAAGAGCTTGTTCACGATCTTGTCTCTGCACTGGAAGAGAGCTCTGAGCAAGCCCGA 214
```

```
Qy     21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln  40
          |||||||||||||||||||||||:::|||||||||:::::::|||||||||||||||||||
Db    215 GGTGGATTTGCTGAAACTGGAGAACATTCTCGAAATCTGTCTTGCCCTCTGAAACGCCAG  274

Qy     41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    275 GCCCGGAAAAGGAGAGGGAGGAAGCGGAGATCCTACAATGTTCACCACCCGTGGGAGACA  334

Qy     61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg  80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    335 GGCCACTGCTTAAGTGAAGGCTCTGATTCTAGTTTAGAAGAACCAAGTAAGGACTATAGA  394

Qy     81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
          |||    |||:::||||||||||||||    ||||||||||||||||||||||||||||||
Db    395 GAGAAGCACAGCAATAATAAAAAGGACCGCAGTGACTCTGATGACCAGATGTTAGTGGCG  454

Qy    101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
          |||||||||||||||||||||||||:::::::::||||||||||||||    |||||||||
Db    455 AAGCGGAGGCCATCTTCAAACCTAAGCAGCAGCGTTCGAGGTAAGCGGCTTCTGTGGCAC  514

Qy    121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
          |||||||||||||||||||||:::::::||||||||||||||||||||||||||||||||
Db    515 GAGTCTGACTTTGCCGTGGACAGCCTTGGGAACAGAACGCTGCGCCGGAGGAGGAAGGTG  574

Qy    141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160
          |||||||||||||||||||||||||||||:::|||:::||||||||||||||||||||
Db    575 AAGCGCATGGCCGTGGATCTCCCGCAGGACGTCTCCAGCAAAAGGACAATGACCCAGCTG  634

Qy    161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
           ||||||||||||||||||||||||||:::||||||||||    ||||||    ||||||
Db    635 CCAGAAGGCTGCAGAGATCAGGACATGGACAATGATAGAGCCAGCCAGTATCCAGAGTTT  694

Qy    181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
          |||:::    |||||||||||||||||||||    ||||||    ||||||||||    |||
Db    695 ACCCGGAAGAAAGTTAAGAAAAGGAAGTTGAAAGGGATTAGGCCAGGACCGAAAACCCAG  754

Qy    201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
          :::|||||||    ||||||||||||||||||||    :::|||    |||||||||:::||||||
Db    755 GAGGAAGGAGGAGTTTTGGAGAGTGAAGAAAGAAGCCAGCCCAACAAGGACAGGATGGAG  814

Qy    221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
             ||||||||||||    |||||||||||    |||||||||||:::||||||||||||
Db    815 TACGAGGAACAGAAAGCCTCGGATGAGCTCAGGAGCGAAAGTGACACCAGCAGTCTCAGC  874

Qy    241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    875 AGCACTGACGCGGGCTTGTTCACCAACGATGAGGGAAGACAAGGTGATGATGAGCAGAGT  934

Qy    261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
          ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db    935 GACTGGTTCTATGAGAAGGAGTCAGGCGGAGCGTGCGGGATTGCTGGAGTCGTGCCCTGG  994

Qy    281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
          |||||||||:::::::|||    |||||||||    |||:::|||||||||||||||||||
Db    995 TGGGAAAAGGATGAGCCAGCAGAGCTGGACACCAACCTGCCCGACCCTGTGTTTGAGAGC
```

1054

Qy   301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320 .
          ||||||:::|||||||||||||||||||||||||||||||| ||| |||||||||||||||||
Db   1055 ATCTTAAGTGGCTCCTTCCCTCTCATGTCCCATCCTGGCAGAGGAGGTTTCCAAGCTAGA
     1114

Qy   321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
          ||||||||||||||||||||| ||||||||||||||||||||||| ||| |||
Db   1115 CTCAGTCGCCTTCATGGAACGCCTTCAAAGAATATTAAAAAGTCTTCAGGGGCTCCACCT
     1174

Qy   341 SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp 360
          ||||||:::||| ||||||||| |||||||||||||||||||||||||||
Db   1175 TCAATGCTGCCTGCTCCTGGCCCTGGCAGTAACAAGAGGATGGTTCACTTCTCCCAGAC
     1234

Qy   361 SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln 380
          :::||| |||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   1235 GCCCATCGCCATGACCACTGGTTTAGCCCTGGGGCTAGGACAGAGCACGGCCAGCATCAG
     1294

Qy   381 LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla 400
          ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||:::::::||||||
Db   1295 CTTCTGAGAGATAACCGAGCGGAAAGAGGACACAAGAAAAGCTGCTCCCTGAAAACAGCC
     1354

Qy   401 SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1355 AGCAGGCAGACAAGCATGCATTTAGGATCCTTGTGCACAGGAGACATCAAAAGGAGAAGG
     1414

Qy   421 LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro 440
          ||||||||||||||||||||||||| |||||| ||||||||||||||||||||||||
Db   1415 AAAGCTGCCCCGTTGCCTGGACCCACTGCAGCAGGAATTGTGGGTGAGAACGCCCAGCCG
     1474

Qy   441 IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer 460
          |||||||||:::||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db   1475 ATCCTAGAGAGCAACATCGGGAACCGCATGCTTCAGAGTATGGGATGGACACCCGGGTCA
     1534

Qy   461 GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys 480
          |||||||||||||||||||||:::||||||||:::|||||||:::||||||:::||||||||||||||
Db   1535 GGCCTCGGGCGAGATGGCAGAGGGATCGCGGAGCCAGTTCAAGCCGTTCAGAGGCCGAAA
     1594

Qy   481 GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla 500
          |||||||||||||||||||||||||||||||||||||||||:::   ::: ||| :::
Db   1595 GGGTTAGGACTTGGATTTCCTCTACCAAAAAGCTCCCCCACCAGCCCTGCCCCCACATCA
     1654

Qy   501 GlyLysSerAla 504
          ||| |||
Db   1655 GGAAACCCTGCC 1666

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 11, 2005, 16:55:46 ; Search time 893 Seconds
                                            (without alignments)
                                            3761.484 Million cell updates/sec

Title: US-09-771-312-2
Perfect score: 2694
Sequence: 1 MEELVHDLVSALEESSEQAR.........GFPLPKSTSATTTPNAGKSA 504

Scoring table: BLOSUM62
                 Xgapop 10.0 , Xgapext 0.5
                 Ygapop 10.0 , Ygapext 0.5
                 Fgapop 6.0 , Fgapext 7.0
                 Delop   6.0 , Delext   7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09771312/runat_01122005_145310_15013/app_query.fasta_1.647
-DB=N_Geneseq -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09771312_@CGN_1_1_727_@runat_01122005_145310_15013 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   N_Geneseq_21:*
                 1: geneseqn1980s:*
                 2: geneseqn1990s:*
                 3: geneseqn2000s:*
                 4: geneseqn2001as:*
                 5: geneseqn2001bs:*
                 6: geneseqn2002as:*
                 7: geneseqn2002bs:*
                 8: geneseqn2003as:*
                 9: geneseqn2003bs:*
                 10: geneseqn2003cs:*
                 11: geneseqn2003ds:*
                 12: geneseqn2004as:*
                 13: geneseqn2004bs:*
                 14: geneseqn2005s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2694	100.0	2338	4	AAH13916	Aah13916 Human cDN
	2	2694	100.0	2338	13	ADR99112	Adr99112 Hypotheti
	3	2694	100.0	2344	6	ABN59701	Abn59701 Novel hum
	4	2694	100.0	2345	5	AAS11663	Aas11663 Prostate
	5	2694	100.0	2583	11	ACN91982	Acn91982 Breast ca
	6	1495.5	55.5	1563	5	AAS72189	Aas72189 DNA encod
	7	1054	39.1	817	4	AAH04312	Aah04312 Human cDN
c	8	967	35.9	553	6	ABT10667	Abt10667 Human bre
c	9	782	29.0	435	13	ACF81403	Acf81403 Human SIR
	10	600	22.3	1656	5	AAS87595	Aas87595 DNA encod
	11	591.5	22.0	1823	6	ABL53700	Abl53700 Human pho
	12	588.5	21.8	2433	6	ABQ61016	Abq61016 14 clone
	13	563.5	20.9	2434	6	ABN59974	Abn59974 Novel hum
	14	530.5	19.7	1474	6	ABN59883	Abn59883 Novel hum
	15	530.5	19.7	2112	4	AAH16534	Aah16534 Human cDN
	16	439	16.3	330	7	ADS72558	Ads72558 Human kid
	17	439	16.3	330	7	ADW41412	Adw41412 cDNA elev
	18	420	15.6	445	4	AAL23787	Aal23787 Human bre
	19	420	15.6	492	11	ACN84952	Acn84952 Breast ca
	20	377	14.0	448	4	AAL14929	Aal14929 Human bre
	21	348.5	12.9	1851	4	AAH13673	Aah13673 Human cDN
	22	321.5	11.9	740	4	AAH03286	Aah03286 Human cDN
	23	320.5	11.9	673	4	AAI21087	Aai21087 Probe #11
	24	320.5	11.9	673	4	ABA66165	Aba66165 Human foe
	25	320.5	11.9	673	4	AAI46350	Aai46350 Probe #15
	26	320.5	11.9	673	4	ABA48281	Aba48281 Human bre
	27	320.5	11.9	673	4	AAK40330	Aak40330 Human bon
	28	320.5	11.9	673	4	AAK14584	Aak14584 Human bra
	29	320.5	11.9	673	4	ABS39901	Abs39901 Human liv
	30	320.5	11.9	673	5	AAI06807	Aai06807 Probe #67
	31	320.5	11.9	673	6	ABS14348	Abs14348 Human gen
	32	320.5	11.9	673	12	ACH91528	Ach91528 Human gen
	33	313	11.6	978	4	AAI11878	Aai11878 Probe #18
	34	313	11.6	978	4	ABA53579	Aba53579 Human foe
	35	313	11.6	978	4	AAI33206	Aai33206 Probe #18
	36	313	11.6	978	4	ABA43160	Aba43160 Human bre
	37	313	11.6	978	4	AAK27304	Aak27304 Human bon
	38	313	11.6	978	4	AAK01845	Aak01845 Human bra
	39	313	11.6	978	4	ABS26879	Abs26879 Human liv
	40	313	11.6	978	5	AAI01816	Aai01816 Probe #18
	41	313	11.6	978	6	ABS01825	Abs01825 Human gen
	42	236.5	8.8	371	4	AAL01647	Aal01647 Human rep
	43	232	8.6	464	9	ACH16845	Ach16845 Human adu
	44	213.5	7.9	747	2	AAZ16358	Aaz16358 Human gen
	45	200.5	7.4	300	2	AAZ14849	Aaz14849 Human gen

ALIGNMENTS

RESULT 1
AAH13916
ID   AAH13916 standard; cDNA; 2338 BP.
XX
AC   AAH13916;
XX
DT   26-JUN-2001 (first entry)
XX

DE   Human cDNA sequence SEQ ID NO:10937.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 10937; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 2338 BP; 739 A; 478 C; 541 G; 580 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.07e-199   Length: 2338
Score: 2694.00   Matches: 504
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0

US-09-771-312-2 (1-504) x AAH13916 (1-2338)

```
Qy      1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    171 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTGGAAGAGAGCTCAGAGCAAGCTCGA 230

Qy     21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 GGTGGATTTGCTGAAACAGGAGACCATTCTCGAAGTATATCTTGCCCTCTGAAACGCCAG 290

Qy     41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 GCAAGGAAAAGGAGAGGGAGAAAACGGAGGTCGTATAATGTGCATCACCCGTGGGAGACT 350

Qy     61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 GGTCACTGCTTAAGTGAAGGCTCTGATTCTAGTTTAGAAGAACCAAGCAAGGACTATAGA 410

Qy     81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 GAGAATCACAATAATAATAAAAAAGATCACAGTGACTCTGATGACCAAATGTTAGTAGCA 470

Qy    101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 AAGCGCAGGCCGTCATCAAACTTAAATAATAATGTTCGAGGGAAAAGACCTCTATGGCAT 530

Qy    121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 GAGTCTGATTTTGCTGTGGACAATGTTGGGAATAGAACTCTGCGCAGGAGGAGAAAGGTA 590

Qy    141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    591 AAACGCATGGCAGTAGATCTCCCACAGGACATCTCTAACAAACGGACAATGACCCAGCCA 650

Qy    161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    651 CCTGAGGGTTGTAGAGATCAGGACATGGACAGTGATAGAGCCTACCAGTATCAAGAATTT 710

Qy    181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    711 ACCAAGAACAAAGTCAAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCAAAAATCCAA 770

Qy    201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    771 GATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGACAAAATGGAA 830

Qy    221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    831 TGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGCAGTCTCAGC 890

Qy    241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    891 AGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGATGAACAGAGT 950

Qy    261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    951 GACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTTGTGCCCTGG 1010

Qy    281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1011 TGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTCTTTGAAAGT 1070
```

```
Qy        301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 ATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTCCAAGCTAGA 1130

Qy        321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1131 CTCAGTCGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAGGGACTCCAACT 1190

Qy        341 SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 TCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATTTTTCCCCGGAT 1250

Qy        361 SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1251 TCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATGACCAGCATCAG 1310

Qy        381 LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1311 CTTCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTCTGTGAGAACAGCC 1370

Qy        401 SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1371 AGCAGGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCAAACGGAGAAGA 1430

Qy        421 LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1431 AAAGCTGCACCTTTGCCTGGACCTACTACTGCAGGATTTGTAGGTGAAAATGCCCAGCCA 1490

Qy        441 IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1491 ATCCTAGAAAATAATATTGGAAACCGAATGCTTCAGAATATGGGCTGGACGCCTGGGTCA 1550

Qy        461 GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1551 GGCCTTGGACGAGATGGCAAGGGGATCTCTGAGCCAATTCAAGCCATGCAGAGGCCAAAG 1610

Qy        481 GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1611 GGATTAGGACTTGGATTTCCTCTACCAAAAAGTACTTCCGCAACTACTACCCCCAATGCA 1670

Qy        501 GlyLysSerAla 504
              ||||||||||||
Db       1671 GGAAAATCCGCC 1682
```

RESULT 2
ADR99112
ID   ADR99112 standard; DNA; 2338 BP.
XX
AC   ADR99112;
XX
DT   02-DEC-2004 (first entry)
XX
DE   Hypothetical protein FLJ10252, coding sequence, SEQ ID 118.
XX
KW   Cytostatic; breast cancer; cancer; human; gene; ds; FLJ10252.
XX
OS   Homo sapiens.
XX
PN   WO2004078035-A2.
XX
PD   16-SEP-2004.
XX

PF   27-FEB-2004; 2004WO-US007268.
XX
PR   28-FEB-2003; 2003US-0450655P.
XX
PA   (FARB) BAYER PHARM CORP.
XX
PI   Eveleigh D, Bigwood D;
XX
DR   WPI; 2004-653556/63.
DR   P-PSDB; ADR99239.
DR   REFSEQ; NM_018040.1.
XX
PT   Diagnosing breast cancer comprises comparing the level of expression of
PT   genes or gene products in a first biological sample taken from a patient
PT   with that in a normal patient sample.
XX
PS   Claim 2; SEQ ID NO 118; 53pp; English.
XX
CC   The present invention relates to a method (M1) for diagnosing breast
CC   cancer in a patient. The method comprises comparing the level of
CC   expression of one or more genes or gene products in a biological sample
CC   from the patient with that in a normal patient sample, where a difference
CC   in the gene expression in the first sample compared to that in the second
CC   sample is a diagnostic of the disease. Also claimed are: method (M2) for
CC   distinguishing between normal and disease tissues; method (M3) for
CC   monitoring the response of a breast cancer patient to treatment with an
CC   anti-cancer agent; method (M4) for identifying a compound for treating
CC   breast cancer; and an array for distinguishing between normal and disease
CC   tissues comprising two or more probes corresponding to genes selected
CC   from ADR98995-ADR99121 or comprising two or more polypeptides selected
CC   from ADR99122-ADR99248. In M1 and M2 the genes are selected from ADR98995
CC   -ADR99121 and the gene products are polypeptides selected from ADR99122-
CC   ADR99248. M1 is useful for diagnosing breast cancer. M2 and the array are
CC   useful for distinguishing between normal and disease tissue. M3 is useful
CC   for monitoring the response of a breast cancer patient to treatment with
CC   an anti-cancer agent. M4 is useful for identifying a compound for
CC   treating breast cancer. Note: The sequence data for this patent did not
CC   form part of the printed specification, but was obtained in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2338 BP; 739 A; 478 C; 541 G; 580 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.07e-199   Length: 2338
Score: 2694.00   Matches: 504
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 13   Gaps: 0

US-09-771-312-2 (1-504) x ADR99112 (1-2338)

Qy   1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   171 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTGGAAGAGAGCTCAGAGCAAGCTCGA 230

Qy   21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   231 GGTGGATTTGCTGAAACAGGAGACCATTCTCGAAGTATATCTTGCCCTCTGAAACGCCAG 290

Qy   41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   291 GCAAGGAAAAGGAGAGGGAGAAACGGAGGTCGTATAATGTGCATCACCCGTGGGAGACT 350

```
Qy    61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   351 GGTCACTGCTTAAGTGAAGGCTCTGATTCTAGTTTAGAAGAACCAAGCAAGGACTATAGA 410

Qy    81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   411 GAGAATCACAATAATAATAAAAAAGATCACAGTGACTCTGATGACCAAATGTTAGTAGCA 470

Qy   101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   471 AAGCGCAGGCCGTCATCAAACTTAAATAATAATGTTCGAGGGAAAAGACCTCTATGGCAT 530

Qy   121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   531 GAGTCTGATTTTGCTGTGGACAATGTTGGGAATAGAACTCTGCGCAGGAGGAGAAAGGTA 590

Qy   141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   591 AAACGCATGGCAGTAGATCTCCCACAGGACATCTCTAACAAACGGACAATGACCCAGCCA 650

Qy   161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   651 CCTGAGGGTTGTAGAGATCAGGACATGGACAGTGATAGAGCCTACCAGTATCAAGAATTT 710

Qy   181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   711 ACCAAGAACAAAGTCAAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCAAAAATCCAA 770

Qy   201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   771 GATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGACAAAATGGAA 830

Qy   221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   831 TGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGCAGTCTCAGC 890

Qy   241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   891 AGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGATGAACAGAGT 950

Qy   261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   951 GACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTTGTGCCCTGG 1010

Qy   281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1011 TGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTCTTTGAAAGT 1070

Qy   301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1071 ATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTCCAAGCTAGA 1130

Qy   321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1131 CTCAGTCGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAGGGACTCCAACT 1190

Qy   341 SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1191 TCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATTTTTCCCCGGAT 1250

Qy   361 SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln 380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1251 TCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATGACCAGCATCAG 1310

Qy    381 LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla 400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1311 CTTCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTCTGTGAGAACAGCC 1370

Qy    401 SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1371 AGCAGGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCAAACGGAGAAGA 1430

Qy    421 LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro 440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1431 AAAGCTGCACCTTTGCCTGGACCTACTACTGCAGGATTTGTAGGTGAAAATGCCCAGCCA 1490

Qy    441 IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer 460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1491 ATCCTAGAAAATAATATTGGAAACCGAATGCTTCAGAATATGGGCTGGACGCCTGGGTCA 1550

Qy    461 GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1551 GGCCTTGGACGAGATGGCAAGGGGATCTCTGAGCCAATTCAAGCCATGCAGAGGCCAAAG 1610

Qy    481 GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla 500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1611 GGATTAGGACTTGGATTTCCTCTACCAAAAAGTACTTCCGCAACTACTACCCCCAATGCA 1670

Qy    501 GlyLysSerAla 504
          ||||||||||||
Db   1671 GGAAAATCCGCC 1682
```

RESULT 3
ABN59701
ID   ABN59701 standard; cDNA; 2344 BP.
XX
AC   ABN59701;
XX
DT   28-JUN-2002 (first entry)
XX
DE   Novel human coding sequence SEQ ID NO: 112.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US026015.
XX
PR   11-SEP-2000; 2000US-00659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   P-PSDB; ABB97288.
XX

PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis.
XX
PS   Claim 1; SEQ ID NO 112; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC   and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC   rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a coding sequence of the
CC   invention
XX
SQ   Sequence 2344 BP; 747 A; 476 C; 541 G; 580 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.08e-199   Length: 2344
Score: 2694.00   Matches: 504
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 6   Gaps: 0

US-09-771-312-2 (1-504) x ABN59701 (1-2344)

```
Qy      1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    170 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTGGAAGAGAGCTCAGAGCAAGCTCGA 229

Qy     21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    230 GGTGGATTTGCTGAAACAGGAGACCATTCTCGAAGTATATCTTGCCCTCTGAAACGCCAG 289

Qy     41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    290 GCAAGGAAAAGGAGAGGGAGAAAACGGAGGTCGTATAATGTGCATCACCCGTGGGAGACT 349

Qy     61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    350 GGTCACTGCTTAAGTGAAGGCTCTGATTCTAGTTTAGAAGAACCAAGCAAGGACTATAGA 409

Qy     81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    410 GAGAATCACAATAATAATAAAAAAGATCACAGTGACTCTGATGACCAAATGTTAGTAGCA 469

Qy    101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    470 AAGCGCAGGCCGTCATCAAACTTAAATAATAATGTTCGAGGGAAAAGACCTCTATGGCAT 529

Qy    121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    530 GAGTCTGATTTTGCTGTGGACAATGTTGGGAATAGAACTCTGCGCAGGAGGAGAAAGGTA 589

Qy    141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    590 AAACGCATGGCAGTAGATCTCCCACAGGACATCTCTAACAAACGGACAATGACCCAGCCA 649

Qy    161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Page 9

```
Db    650 CCTGAGGGTTGTAGAGATCAGGACATGGACAGTGATAGAGCCTACCAGTATCAAGAATTT 709

Qy    181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    710 ACCAAGAACAAAGTCAAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCAAAAATCCAA 769

Qy    201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    770 GATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGACAAAATGGAA 829

Qy    221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    830 TGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGCAGTCTCAGC 889

Qy    241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    890 AGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGATGAACAGAGT 949

Qy    261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    950 GACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTTGTGCCCTGG 1009

Qy    281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1010 TGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTCTTTGAAAGT 1069

Qy    301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 ATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTCCAAGCTAGA 1129

Qy    321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130 CTCAGTCGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAGGGACTCCAACT 1189

Qy    341 SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190 TCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATTTTTCCCCGGAT 1249

Qy    361 SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln 380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250 TCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATGACCAGCATCAG 1309

Qy    381 LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla 400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310 CTTCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTCTGTGAGAACAGCC 1369

Qy    401 SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1370 AGCAGGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCAAACGGAGAAGA 1429

Qy    421 LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro 440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1430 AAAGCTGCACCTTTGCCTGGACCTACTACTGCAGGATTTGTAGGTGAAAATGCCCAGCCA 1489

Qy    441 IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer 460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1490 ATCCTAGAAAATAATATTGGAAACCGAATGCTTCAGAATATGGGCTGGACGCCTGGGTCA 1549

Qy    461 GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1550 GGCCTTGGACGAGATGGCAAGGGGATCTCTGAGCCAATTCAAGCCATGCAGAGGCCAAAG 1609

Qy    481 GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla 500
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1610  GGATTAGGACTTGGATTTCCTCTACCAAAAAGTACTTCCGCAACTACTACCCCCAATGCA 1669

Qy     501  GlyLysSerAla 504
            ||||||||||||
Db    1670  GGAAAATCCGCC 1681
```

RESULT 4
AAS11663
ID   AAS11663 standard; cDNA; 2345 BP.
XX
AC   AAS11663;
XX
DT   24-OCT-2001 (first entry)
XX
DE   Prostate and testis-related gene 84P2A9 cDNA.
XX
KW   84P2A9; PCR primer; DNA adaptor; prostate; testis; tissue; cancer; ss;
KW   leukaemia; tumour; kidney; brain; bone; skin; ovary; breast; pancreas;
KW   colon; lung; cytostatic; gene therapy; antibody therapy; ribozyme;
KW   single chain monoclonal antibody; serum; blood; urine.
XX
OS   Homo sapiens.
XX
PN   WO200155391-A2.
XX
PD   02-AUG-2001.
XX
PF   26-JAN-2001; 2001WO-US002651.
XX
PR   26-JAN-2000; 2000US-0178560P.
XX
PA   (UROG-) UROGENESYS INC.
XX
PI   Jakobovits A, Afar DEH, Challita-Eid PM, Levin E, Mitchell SC;
PI   Hubert RS;
XX
DR   WPI; 2001-502631/55.
DR   P-PSDB; AAU06524.
XX
PT   New 84P2A9 gene and its encoded protein, useful for diagnosing and
PT   treating cancer, e.g. leukemia and cancer of the prostate, testis,
PT   kidney, brain or bone, or for eliciting an immune response.
XX
PS   Claim 1; Fig 2; 149pp; English.
XX
CC   The nucleic acid sequences represent the 84P2A9 gene and the primers and
CC   adaptors used to amplify 84P2A9 DNA. 84P2A9 exhibits prostate and testis
CC   specific expression in normal adult tissue, but it is also aberrantly
CC   expressed in many cancers including leukaemia and tumours of the
CC   prostate, testis, kidney, brain, bone, skin, ovary, breast, pancreas,
CC   colon and lung. The 84P2A9 polynucleotide, its related protein and also
CC   peptide fragments of the protein are therefore useful for diagnosing and
CC   treating cancer. A vector comprising a polynucleotide which encodes a
CC   single chain monoclonal antibody, that immunospecifically binds to an
CC   84P2A9-related protein, and a ribozyme capable of cleaving a
CC   polynucleotide having the 84P2A9 coding sequence, are both useful in the
CC   preparation of a composition for treating a patient with a cancer that
CC   expresses 84P2A9. The sequences can be used in diagnostic methods to
CC   monitor the level of 84P2A9 gene products in serum, blood, urine and
CC   tissue and to thereby detect the presence of cancerous cells
XX
SQ   Sequence 2345 BP; 750 A; 476 C; 542 G; 577 T; 0 U; 0 Other;

                              Page 11

Alignment Scores:
Pred. No.:   2.08e-199   Length: 2345
Score: 2694.00   Matches: 504
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 5   Gaps: 0

US-09-771-312-2 (1-504) x AAS11663 (1-2345)

```
Qy      1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    163 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTGGAAGAGAGCTCAGAGCAAGCTCGA 222

Qy     21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    223 GGTGGATTTGCTGAAACAGGAGACCATTCTCGAAGTATATCTTGCCCTCTGAAACGCCAG 282

Qy     41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    283 GCAAGGAAAAGGAGAGGGAGAAAACGGAGGTCGTATAATGTGCATCACCCGTGGGAGACT 342

Qy     61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    343 GGTCACTGCTTAAGTGAAGGCTCTGATTCTAGTTTAGAAGAACCAAGCAAGGACTATAGA 402

Qy     81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    403 GAGAATCACAATAATAATAAAAAAGATCACAGTGACTCTGATGACCAAATGTTAGTAGCA 462

Qy    101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    463 AAGCGCAGGCCGTCATCAAACTTAAATAATAATGTTCGAGGGAAAAGACCTCTATGGCAT 522

Qy    121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    523 GAGTCTGATTTTGCTGTGGACAATGTTGGGAATAGAACTCTGCGCAGGAGGAGAAAGGTA 582

Qy    141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    583 AAACGCATGGCAGTAGATCTCCCACAGGACATCTCTAACAAACGGACAATGACCCAGCCA 642

Qy    161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    643 CCTGAGGGTTGTAGAGATCAGGACATGGACAGTGATAGAGCCTACCAGTATCAAGAATTT 702

Qy    181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    703 ACCAAGAACAAAGTCAAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCAAAAATCCAA 762

Qy    201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    763 GATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGACAAAATGGAA 822

Qy    221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    823 TGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGCAGTCTCAGC 882

Qy    241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    883 AGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGATGAACAGAGT 942
```

```
Qy     261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     943 GACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTTGTGCCCTGG 1002

Qy     281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1003 TGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTCTTTGAAAGT 1062

Qy     301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1063 ATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTCCAAGCTAGA 1122

Qy     321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1123 CTCAGTCGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAGGGACTCCAACT 1182

Qy     341 SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1183 TCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATTTTTCCCCGGAT 1242

Qy     361 SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln 380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1243 TCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATGACCAGCATCAG 1302

Qy     381 LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla 400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1303 CTTCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTCTGTGAGAACAGCC 1362

Qy     401 SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1363 AGCAGGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCAAACGGAGAAGA 1422

Qy     421 LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro 440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1423 AAAGCTGCACCTTTGCCTGGACCTACTACTGCAGGATTTGTAGGTGAAAATGCCCAGCCA 1482

Qy     441 IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer 460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1483 ATCCTAGAAAATAATATTGGAAACCGAATGCTTCAGAATATGGGCTGGACGCCTGGGTCA 1542

Qy     461 GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1543 GGCCTTGGACGAGATGGCAAGGGGATCTCTGAGCCAATTCAAGCCATGCAGAGGCCAAAG 1602

Qy     481 GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla 500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1603 GGATTAGGACTTGGATTTCCTCTACCAAAAAGTACTTCCGCAACTACTACCCCCAATGCA 1662

Qy     501 GlyLysSerAla 504
           ||||||||||||
Db    1663 GGAAAATCCGCC 1674
```

RESULT 5
ACN91982
ID   ACN91982 standard; DNA; 2583 BP.
XX
AC   ACN91982;
XX
DT   02-DEC-2004 (first entry)
XX
DE   Breast cancer related marker, seq id 13132.
XX
KW   Cancer; breast; tumour; cytostatic; marker; detection; therapy; ds.

```
XX
OS    Homo sapiens.
XX
PN    US2003099974-A1.
XX
PD    29-MAY-2003.
XX
PF    18-JUL-2002; 2002US-00198846.
XX
PR    18-JUL-2001; 2001US-0306220P.
XX
PA    (MILL-) MILLENNIUM PHARM INC.
XX
PI    Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR    WPI; 2003-787014/74.
XX
PT    Novel isolated polypeptide associated with breast cancer, useful for
PT    detecting presence of polypeptide in sample, as a marker for breast
PT    cancer.
XX
PS    Disclosure; SEQ ID NO 13132; 36pp; English.
XX
CC    The invention relates to an isolated polypeptide (I) associated with
CC    breast cancer which is encoded by a nucleic acid molecule comprising a
CC    nucleotide sequence (S1). Further disclosed is an antibody that binds to
CC    the polypeptide of the invention. The activity of the polypeptide of the
CC    invention may be described as cytostatic. The antibody is useful for
CC    detecting the presence of (I) in a sample. Nucleic acid molecules of the
CC    invention are useful in the detection of breast tumours. (I) is useful as
CC    a marker for breast cancer and in breast cancer therapy. Sequences given
CC    in records ACN78851-ACN92934 represent nucleic acid markers associated
CC    with breast cancer. Note: The sequence listing does not form part of the
CC    specification but may be obtained in electronic format from the USPTO web
CC    site at seqdata.uspto.gov/sequence.html?DocID=20030099974
XX
SQ    Sequence 2583 BP; 813 A; 519 C; 575 G; 664 T; 0 U; 12 Other;
```

Alignment Scores:

Pred. No.:	2.35e-199	Length:	2583
Score:	2694.00	Matches:	504
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	11	Gaps:	0

US-09-771-312-2 (1-504) x ACN91982 (1-2583)

```
Qy      1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    199 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTGGAAGAGAGCTCAGAGCAAGCTCGA 258

Qy     21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    259 GGTGGATTTGCTGAAACAGGAGACCATTCTCGAAGTATATCTTGCCCTCTGAAACGCCAG 318

Qy     41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 GCAAGGAAAAGGAGAGGGAGAAAACGGAGGTCGTATAATGTGCATCACCCGTGGGAGACT 378

Qy     61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    379 GGTCACTGCTTAAGTGAAGGCTCTGATTCTAGTTTAGAAGAACCAAGCAAGGACTATAGA 438
```

```
Qy     81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    439 GAGAATCACAATAATAATAAAAAAGATCACAGTGACTCTGATGACCAAATGTTAGTAGCA 498

Qy    101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    499 AAGCGCAGGCCGTCATCAAACTTAAATAATAATGTTCGAGGGAAAAGACCTCTATGGCAT 558

Qy    121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    559 GAGTCTGATTTTGCTGTGGACAATGTTGGGAATAGAACTCTGCGCAGGAGGAGAAAGGTA 618

Qy    141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    619 AAACGCATGGCAGTAGATCTCCCACAGGACATCTCTAACAAACGGACAATGACCCAGCCA 678

Qy    161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    679 CCTGAGGGTTGTAGAGATCAGGACATGGACAGTGATAGAGCCTACCAGTATCAAGAATTT 738

Qy    181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    739 ACCAAGAACAAAGTCAAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCAAAAATCCAA 798

Qy    201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    799 GATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGACAAAATGGAA 858

Qy    221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    859 TGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGCAGTCTCAGC 918

Qy    241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    919 AGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGATGAACAGAGT 978

Qy    261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    979 GACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTTGTGCCCTGG 1038

Qy    281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1039 TGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTCTTTGAAAGT 1098

Qy    301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1099 ATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTCCAAGCTAGA 1158

Qy    321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1159 CTCAGTCGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAGGGACTCCAACT 1218

Qy    341 SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1219 TCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATTTTTCCCCGGAT 1278

Qy    361 SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1279 TCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATGACCAGCATCAG 1338

Qy    381 LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    1339 CTTCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTCTGTGAGAACAGCC 1398

Qy     401 SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1399 AGCAGGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCAAACGGAGAAGA 1458

Qy     421 LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro 440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1459 AAAGCTGCACCTTTGCCTGGACCTACTACTGCAGGATTTGTAGGTGAAAATGCCCAGCCA 1518

Qy     441 IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer 460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1519 ATCCTAGAAAATAATATTGGAAACCGAATGCTTCAGAATATGGGCTGGACGCCTGGGTCA 1578

Qy     461 GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1579 GGCCTTGGACGAGATGGCAAGGGGATCTCTGAGCCAATTCAAGCCATGCAGAGGCCAAAG 1638

Qy     481 GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla 500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1639 GGATTAGGACTTGGATTTCCTCTACCAAAAAGTACTTCCGCAACTACTACCCCAATGCA 1698

Qy     501 GlyLysSerAla 504
           ||||||||||||
Db    1699 GGAAAATCCGCC 1710
```

RESULT 6
AAS72189
ID   AAS72189 standard; cDNA; 1563 BP.
XX
AC   AAS72189;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #7993.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG08002.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 7993; 103pp; English.

```
XX
CC    The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC    sequences. (I) is useful as hybridisation probes, polymerase chain
CC    reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC    and in recombinant production of (II). The polynucleotides are also used
CC    in diagnostics as expressed sequence tags for identifying expressed
CC    genes. (I) is useful in gene therapy techniques to restore normal
CC    activity of (II) or to treat disease states involving (II). (II) is
CC    useful for generating antibodies against it, detecting or quantitating a
CC    polypeptide in tissue, as molecular weight markers and as a food
CC    supplement. (II) and its binding partners are useful in medical imaging
CC    of sites expressing (II). (I) and (II) are useful for treating disorders
CC    involving aberrant protein expression or biological activity. The
CC    polypeptide and polynucleotide sequences have applications in
CC    diagnostics, forensics, gene mapping, identification of mutations
CC    responsible for genetic disorders or other traits to assess biodiversity
CC    and to produce other types of data and products dependent on DNA and
CC    amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC    coding sequences of the invention. Note: The sequence data for this
CC    patent did not appear in the printed specification, but was obtained in
CC    electronic format directly from WIPO at
CC    ftp.wipo.int/pub/published_pct_sequences
XX
SQ    Sequence 1563 BP; 486 A; 326 C; 395 G; 356 T; 0 U; 0 Other;
```

Alignment Scores:

Pred. No.:	1.55e-106	Length:	1563
Score:	1495.50	Matches:	312
Percent Similarity:	69.80%	Conservative:	7
Best Local Similarity:	68.27%	Mismatches:	21
Query Match:	55.51%	Indels:	117
DB:	5	Gaps:	2

US-09-771-312-2 (1-504) x AAS72189 (1-1563)

```
Qy      1 MetGluGluLeuValHisAspLeuValSerAlaLeuGlu---GluSerSerGluGlnAla 19
          |||||||||||||||||||||||||||||||||||||||   |||         :::
Db    155 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTGGAAAGAGAGCTCCAGAGCAAGCCT 214

Qy     20 ArgGlyGlyPheAlaGluThrGlyAsp-HisSer-ArgSerIleSerCysProLeuLysA 39
          |||||||||||||||||||||   ||||||  ||||||  ||||||||||||||||||||||||
Db    215 CGAGGTGGATTTGCTGAACCAGGAGACCCATTCTCCGAAGTATATCTTGCCCTCTGAAAC 274

Qy     39 rgGlnAlaArgLysArgArgGlyArgLysArgArg-SerTyrAsnValHisHisProTrp 58
          ||       ||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    275 GCCCAGCAAGGAAAAGGAGAGGGAGAAAACGGAGGTTCGTATAATGTGCATCACCCGTGG 334

Qy     59 Glu-ThrGlyHisCysLeu--SerGluGlySerAspSerSerLeuGluGluProSerLys 77
          ||| |||||||||   ||| |||||||||||||||||||
Db    335 GAGGACTGGTCACTGGCTTAAAGTGAAGGCTCTGATTCTAGT----------------- 376

Qy     78 AspTyrArgGluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMet 97

Db    376 ------------------------------------------------------------ 376

Qy     98 LeuValAlaLysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgPro 117

Db    376 ------------------------------------------------------------ 376

Qy    118 LeuTrpHisGluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArg 137

Db    376 ------------------------------------------------------------ 376
```

```
Qy   138 ArgLysValLysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMet 157

Db   376 ------------------------------------------------------------ 376


Qy   158 ThrGlnProProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyr 177

Db   376 ------------------------------------------------------------ 376


Qy   178 GlnGluPheThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyPro 197
             |||   :::   |||   |||||||||||||||||||||||||||||||||||||||
Db   377 ------TTTAGAAGAACAAAGTCAAAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCA 430


Qy   198 LysIleGlnAspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAsp 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   431 AAAATCCAAGATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGAC 490


Qy   218 LysMetGluCysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSer 237
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   491 AAAATGGAATGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGC 550


Qy   238 SerLeuSerSerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAsp 257
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   551 AGTCTCAGCAGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGAT 610


Qy   258 GluGlnSerAspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyVal 277
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   611 GAACAGAGTGACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTT 670


Qy   278 ValProTrpTrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProVal 297
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   671 GTGCCCTGGTGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTC 730


Qy   298 PheGluSerIleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGly-Ph 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   731 TTTGAAAGTATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTT 790


Qy   317 eGlnAlaArgLeuSerArg-LeuHisGlyMetSerSerLysAsnIleLysLysSerGlyG 337
         ||||     ||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   791 CCAACTAAGACTCAGTCGGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAG 850


Qy   337 lyThrProThrSerMetValProIleProGlyProValGlyAsnLysArgMetValHisP 357
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   851 GGACTCCAACTTCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATT 910


Qy   357 heSerProAspSerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisA 377
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   911 TTTCCCCGGATTCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATG 970


Qy   377 spGlnHisGlnLeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerV 397
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   971 ACCAGCATCAGCTTCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTCTG 1030


Qy   397 alArgThrAlaSerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleL 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1031 TGAGAACAGCCAGCAGGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCA 1090


Qy   417 ysArgArgArgLysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluA 437
         |||||||||||||||||||||||||||||||||||||||||||     :::
Db  1091 AACGGAGAAGAAAAGCTGCACCTTTGCCTGGACCTACTACTGCAGATTATTTCTCCCCCA 1150


Qy   437 snAlaGlnProIleLeuGluAsnAsnIleGlyAsn 448
         :::|||:::::::       |||:::
Db  1151 TTCCCAAGCCAGTTATAGTAAAAGAATGTGGAAGT 1185
```

```
                         GenCore version 5.1.6
                  Copyright (c) 1993 - 2005  Compugen Ltd.


OM protein - protein search, using sw model

Run on:          December  4, 2005, 10:07:40 ; Search time 41 Seconds
                                            (without alignments)
                                            1182.762 Million cell updates/sec

Title:           US-09-771-312-2
Perfect score:   2694
Sequence:        1 MEELVHDLVSALEESSEQAR.........GFPLPKSTSATTTPNAGKSA 504

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters:     283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_80:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                      %
Result            Query
   No.    Score   Match Length DB   ID                 Description
   ----------------------------------------------------------------------------
     1    177.5    6.6    1105  2   T47582             hypothetical prote
     2    167.5    6.2     767  2   S63182             hypothetical prote
     3    156      5.8     542  2   T46464             hypothetical prote
     4    149      5.5     695  2   T40168             hypothetical prote
     5    148.5    5.5    1403  1   A47328             natural killer cel
     6    146.5    5.4     669  2   T28754             hypothetical prote
     7    143      5.3    1577  2   T19722             hypothetical prote
     8    143      5.3    3498  2   T22330             hypothetical prote
     9    138.5    5.1     368  2   G88636             protein W09G12.7 [
    10    135.5    5.0     643  2   A96636             unknown protein, 7
    11    134.5    5.0     699  2   I38073             nucleolar phosphop
    12    134.5    5.0     896  2   D96556             hypothetical prote
    13    133.5    5.0    1672  2   T46237             hypothetical prote
```

```
14        133     4.9      705    2   D88536              acidic protein - C
15        133     4.9      705    2   S27786              acidic protein - C
16        133     4.9      943    2   A42681              centromere protein
17        131.5   4.9      425    2   S55147              hypothetical prote
18        130     4.8      608    2   T02299              hypothetical prote
19        130     4.8      679    2   S48437              hypothetical prote
20        129.5   4.8     2526    2   T20531              hypothetical prote
21        129.5   4.8     2722    2   T20532              hypothetical prote
22        129.5   4.8     2738    2   E88320              protein F07A11.6 [
23        128.5   4.8      543    2   T27190              hypothetical prote
24        128.5   4.8      552    2   T27191              hypothetical prote
25        128.5   4.8      954    2   E86174              protein F19P19.26
26        127.5   4.7      493    2   T02376              hypothetical prote
27        127.5   4.7      539    2   T15256              hypothetical prote
28        127     4.7      763    2   T08929              hypothetical prote
29        127     4.7      786    2   T33856              hypothetical prote
30        127     4.7      845    2   A45669              neurofilament trip
31        127     4.7      963    2   T04002              hypothetical prote
32        126.5   4.7      390    2   T34137              hypothetical prote
33        126     4.7      598    2   B40713              cylicin I - human
34        126     4.7     1032    2   A57514              RNA helicase HEL11
35        125     4.6     1274    2   A89959              hypothetical prote
36        124.5   4.6      817    2   S53919              hypothetical prote
37        124     4.6      775    2   T21259              hypothetical prote
38        124     4.6     1166    2   H86341              hypothetical prote
39        123.5   4.6      849    2   E86306              Similar to tufteli
40        123     4.6      529    2   T50609              hypothetical prote
41        122     4.5      581    2   T22455              hypothetical prote
42        122     4.5      611    2   T22456              hypothetical prote
43        122     4.5      971    2   T24866              hypothetical prote
44        122     4.5     1230    2   T22458              hypothetical prote
45        121     4.5     4910    2   S64942              probable membrane
```

ALIGNMENTS

RESULT 1
T47582
hypothetical protein F24B22.190 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C;Accession: T47582
R;Bloecker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat, M.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23016
A;Accession: T47582
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1105 <BLO>
A;Cross-references: UNIPROT:Q9M383; UNIPARC:UPI00000A410D; EMBL:AL132957
A;Experimental source: cultivar Columbia; BAC clone F24B22
C;Genetics:
A;Map position: 3
A;Introns: 35/3; 56/2; 294/3; 318/3; 349/3; 376/2; 426/3; 455/1; 485/3; 508/3;
568/3; 633/1; 662/3; 681/3; 710/2; 981/1; 1043/3

A;Note: F24B22.190

```
    Query Match                6.6%;  Score 177.5;  DB 2;  Length 1105;
    Best Local Similarity     23.0%;  Pred. No. 0.00075;
    Matches   90;  Conservative   48;  Mismatches  146;  Indels  107;  Gaps   14;


Qy       158 TQPPEGCRDQDMDSDRAYQYQEFTKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQT--- 214
             ||    | ||: :| :  : |  :  | |  : |:|       :    ||   :  |
Db       722 TQQYVPCPDQNNES-KVTENQPDSAKKEKSSQQKVIISAATTPNVEKVLSLPDAVQAAAA 780


Qy       215 ---NKDKMECEEQKVSDELMSESDSSSLSSTDAGLFTNDEGRQGDDEQSDWFYEKESGGA 271
                :| | | |   | |:  | :| |:|                        :|:
Db       781 AAIASEKREKERVK---EIKLASKTSLLAS----------------------KKKMSNV 814


Qy       272 CGITGVVPWWEKEDPTELDKNVPDPVFESILTGSFPLMSHPSRRGF------------- 317
               :   | ::   |:: :   |          | : | :|  :|   |
Db       815 -----LTMWKQRSHETQIQRPSPS------LGDNPPTVSAEARSSFSTGQSMGKLKSDVI 863


Qy       318 --QARLSRLHGMS------SKNIKKSGGT----------------PTSMVPIPG--PVG 350
               : | :  ||:|       | :  :|||                 :: | :|    |
Db       864 IAKERSTSNHGVSALTTAESSSSSTTGGTLMGVMRGSFGGTLGGASSSASVQMPPILPSA 923


Qy       351 NKRMVHFSPDSHHHDHWFSPGARTEHDQHQL-LRDNRAERGHKKNCSVRTASR--QTSMH 407
               :   |   |             | :|| |  ||  ||| :    |  : :     :|
Db       924 SPASVSVSGSGRRRFSETPTAGPTHREQPQTSYRDRAAERRNLYGSSTSSGNDVIDSSED 983


Qy       408 LGSLCTGDIKRRRKAAPLPGPTTAGFVG-------------ENAQPILENNIGNRMLQNM 454
             | |       |: : | | |  | |                 : | |:|:||||:||
Db       984 LMGL------RKGSSDPTPFPPGVGGRGITTSTEVSSFDVITEERAIDESNVGNRMLRNM 1037


Qy       455 GWTPGSGLGRDGKGISEPIQAMQRPKGLGLG 485
             ||  |||||:|| |: ||:||   :   |||
Db      1038 GWHEGSGLGKDGSGMKEPVQAQGVDRRAGLG 1068
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005   Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 11, 2005, 17:38:44 ; Search time 5807 Seconds
                                    (without alignments)
                                    4060.735 Million cell updates/sec

Title: US-09-771-312-2
Perfect score: 2694
Sequence: 1 MEELVHDLVSALEESSEQAR.........GFPLPKSTSATTTPNAGKSA 504

Scoring table: BLOSUM62
                 Xgapop 10.0 , Xgapext 0.5
                 Ygapop 10.0 , Ygapext 0.5
                 Fgapop 6.0 , Fgapext 7.0
                 Delop   6.0 , Delext   7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-
Q=/cgn2_1/USPTO_spool/US09771312/runat_01122005_145312_15071/app_query.fasta_1.6
47
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09771312_@CGN_1_1_5315_@runat_01122005_145312_15071 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
                 1: gb_est1:*
                 2: gb_est2:*
                 3: gb_est3:*
                 4: gb_htc:*
                 5: gb_est4:*
                 6: gb_est5:*
                 7: gb_est6:*
                 8: gb_est7:*
                 9: gb_gss1:*
                 10: gb_gss2:*

11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2694	100.0	1587	11	DQ052881	DQ052881 Homo sapi
	2	2340	86.9	1587	11	DQ052882	DQ052882 Pan trogl
	3	2301	85.4	3844	4	AK030026	AK030026 Mus muscu
	4	2275	84.4	2746	4	AK029990	AK029990 Mus muscu
	5	2089.5	77.6	2234	4	AK053781	AK053781 Mus muscu
	6	1614	59.9	1961	4	AK032734	AK032734 Mus muscu
	7	1541	57.2	1565	4	AK017975	AK017975 Mus muscu
	8	1389	51.6	1099	1	AL539463	AL539463 AL539463
	9	1338	49.7	770	7	CN365216	CN365216 170004247
	10	1303	48.4	3704	4	AK083471	AK083471 Mus muscu
	11	1209	44.9	968	2	BG335967	BG335967 602404712
	12	1200	44.5	923	5	BX380890	BX380890 BX380890
	13	1179	43.8	790	5	BX415500	BX415500 BX415500
	14	1129.5	41.9	845	5	BW972661	BW972661 BW972661
	15	1108	41.1	756	7	CN457551	CN457551 UI-M-HN0-
	16	1100	40.8	782	5	BX355142	BX355142 BX355142
	17	1054	39.1	817	1	AU120500	AU120500 AU120500
	18	1040	38.6	593	5	BX506191	BX506191 DKFZp686P
c	19	1008	37.4	676	3	BM683630	BM683630 UI-E-EJ1-
	20	1005.5	37.3	630	8	DN532759	DN532759 1360834 M
c	21	998	37.0	659	6	CB046496	CB046496 NISC_gf04
	22	997.5	37.0	716	2	BB654618	BB654618 BB654618
c	23	988	36.7	555	5	BU952562	BU952562 io76b05.x
c	24	988	36.7	560	7	CK820603	CK820603 id99a12.y
c	25	988	36.7	725	1	AJ731509	AJ731509 AJ731509
c	26	984	36.5	572	3	BI791523	BI791523 id99a12.x
	27	983.5	36.5	760	7	CK460603	CK460603 930893 MA
	28	981	36.4	813	7	CK633792	CK633792 UI-M-HN0-
	29	979	36.3	702	3	BI766808	BI766808 603056721
c	30	974	36.2	547	1	AI694727	AI694727 we42c09.x
	31	972	36.1	880	7	CK600754	CK600754 AGENCOURT
	32	968	35.9	678	5	BX955296	BX955296 DKFZp781N
c	33	967	35.9	553	1	AI949698	AI949698 wq13h04.x
c	34	965	35.8	542	2	BE501500	BE501500 hw32h06.x
	35	941	34.9	667	3	BM929686	BM929686 UI-E-EJ1-
	36	941	34.9	684	8	CX165742	CX165742 HESC2_32_
	37	919	34.1	638	6	CB554977	CB554977 MMSP0041_
	38	915	34.0	778	7	CO431345	CO431345 UI-M-HX0-
	39	882.5	32.8	791	5	BU480355	BU480355 603842912
	40	881	32.7	675	3	BI560845	BI560845 603254011
	41	872.5	32.4	712	3	BP148863	BP148863 BP148863
	42	860	31.9	717	7	CN535149	CN535149 UI-M-HS0-
c	43	858	31.8	489	1	AI436121	AI436121 ti15f01.y
	44	856	31.8	647	5	BU951125	BU951125 io76b05.y
c	45	837	31.1	536	2	BG063093	BG063093 H3002A11-

RESULT 1
DQ052881
LOCUS   DQ052881   1587 bp   DNA   linear   GSS 02-JUN-2005
DEFINITION   Homo sapiens FLJ10252 gene, VIRTUAL TRANSCRIPT, partial sequence,
               genomic survey sequence.
ACCESSION   DQ052881
VERSION   DQ052881.1 GI:66898828
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE   1 (bases 1 to 1587)
  AUTHORS   Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
               Hubisz,M.J., Fledel-Alon,A., Tanenbaum,D.M., Civello,D.,
               White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
  TITLE   A Scan for Positively Selected Genes in the Genomes of Humans and
               Chimpanzees
  JOURNAL   (er) PLoS Biol. 3 (6), E170 (2005)
  PUBMED   15869325
REFERENCE   2 (bases 1 to 1587)
  AUTHORS   Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
               Hubisz,M.J., Fledel-Alon,A., Tanenbaum,D.M., Civello,D.,
               White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (05-MAY-2005) Celera Genomics, 45 West Gude Drive,
               Rockville, MD 20850, USA
COMMENT   This sequence was made by sequencing genomic exons and ordering
               them based on alignment. Translation starts at the beginning of
               alignment.
FEATURES   Location/Qualifiers
     source   1. .1587
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
     gene   <1. .>1587
                     /gene="FLJ10252"
                     /locus_tag="HC13411"
ORIGIN

Alignment Scores:
Pred. No.:   2.81e-236   Length: 1587
Score: 2694.00   Matches: 504
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 11   Gaps: 0

US-09-771-312-2 (1-504) x DQ052881 (1-1587)

Qy   1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db        73 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTGGAAGAGAGCTCAGAGCAAGCTCGA 132

Qy        21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       133 GGTGGATTTGCTGAAACAGGAGACCATTCTCGAAGTATATCTTGCCCTCTGAAACGCCAG 192

Qy        41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       193 GCAAGGAAAAGGAGAGGGAGAAAACGGAGGTCGTATAATGTGCATCACCCGTGGGAGACT 252

Qy        61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       253 GGTCACTGCTTAAGTGAAGGCTCTGATTCTAGTTTAGAAGAACCAAGCAAGGACTATAGA 312

Qy        81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       313 GAGAATCACAATAATAATAAAAAGATCACAGTGACTCTGATGACCAAATGTTAGTAGCA 372

Qy       101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       373 AAGCGCAGGCCGTCATCAAACTTAAATAATAATGTTCGAGGGAAAAGACCTCTATGGCAT 432

Qy       121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       433 GAGTCTGATTTTGCTGTGGACAATGTTGGGAATAGAACTCTGCGCAGGAGGAGAAAGGTA 492

Qy       141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       493 AAACGCATGGCAGTAGATCTCCCACAGGACATCTCTAACAAACGGACAATGACCCAGCCA 552

Qy       161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       553 CCTGAGGGTTGTAGAGATCAGGACATGGACAGTGATAGAGCCTACCAGTATCAAGAATTT 612

Qy       181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       613 ACCAAGAACAAAGTCAAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCAAAAATCCAA 672

Qy       201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       673 GATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGACAAAATGGAA 732

Qy       221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       733 TGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGCAGTCTCAGC 792

Qy       241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       793 AGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGATGAACAGAGT 852

Qy       261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       853 GACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTTGTGCCCTGG 912

Qy       281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       913 TGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTCTTTGAAAGT 972
```

```
Qy        301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 ATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTCCAAGCTAGA 1032

Qy        321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 CTCAGTCGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAGGGACTCCAACT 1092

Qy        341 SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 TCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATTTTTCCCCGGAT 1152

Qy        361 SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 TCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATGACCAGCATCAG 1212

Qy        381 LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 CTTCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTCTGTGAGAACAGCC 1272

Qy        401 SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1273 AGCAGGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCAAACGGAGAAGA 1332

Qy        421 LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1333 AAAGCTGCACCTTTGCCTGGACCTACTACTGCAGGATTTGTAGGTGAAAATGCCCAGCCA 1392

Qy        441 IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1393 ATCCTAGAAAATAATATTGGAAACCGAATGCTTCAGAATATGGGCTGGACGCCTGGGTCA 1452

Qy        461 GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1453 GGCCTTGGACGAGATGGCAAGGGGATCTCTGAGCCAATTCAAGCCATGCAGAGGCCAAAG 1512

Qy        481 GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1513 GGATTAGGACTTGGATTTCCTCTACCAAAAAGTACTTCCGCAACTACTACCCCCAATGCA 1572

Qy        501 GlyLysSerAla 504
              ||||||||||||
Db       1573 GGAAAATCCGCC 1584
```

RESULT 2
DQ052882
LOCUS   DQ052882   1587 bp   DNA   linear   GSS 02-JUN-2005
DEFINITION   Pan troglodytes FLJ10252 gene, VIRTUAL TRANSCRIPT, partial
                sequence, genomic survey sequence.
ACCESSION   DQ052882
VERSION   DQ052882.1 GI:66898829
KEYWORDS   GSS.
SOURCE   Pan troglodytes (chimpanzee)
  ORGANISM   Pan troglodytes
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

  COMMENT   This sequence was made by sequencing genomic exons and ordering
            them based on alignment. Translation starts at the beginning of
            alignment.
FEATURES   Location/Qualifiers
     source   1. .1587
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
     gene   <1._.>1587
                     /gene="FLJ10252"
                     /locus_tag="HC13411"
ORIGIN

Alignment Scores:
Pred. No.:   8.94e-204   Length: 1587
Score: 2340.00   Matches: 445
Percent Similarity: 88.29%   Conservative: 0
Best Local Similarity: 88.29%   Mismatches: 59
Query Match: 86.86%   Indels: 0
DB: 11   Gaps: 0

US-09-771-312-2 (1-504) x DQ052882 (1-1587)

Qy   1 MetGluGluLeuValHisAspLeuValSerAlaLeuGluGluSerSerGluGlnAlaArg 20
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73 ATGGAGGAGCTGGTTCATGACCTTGTCTCAGCATTAGAAGAGAGCTCAGAGCAAGCTCGA 132

Qy   21 GlyGlyPheAlaGluThrGlyAspHisSerArgSerIleSerCysProLeuLysArgGln 40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   133 GGTGGATTTGCTGAAACAGGAGACCATTCTCGAAGTATATCTTGCCCTCTGAAACGCCAG 192

Qy   41 AlaArgLysArgArgGlyArgLysArgArgSerTyrAsnValHisHisProTrpGluThr 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   193 GCAAGGAAAAGGAGAGGGAGAAAACGGAGGTCGTATAATGTGCATCACCCGTGGGAGACT 252

Qy   61 GlyHisCysLeuSerGluGlySerAspSerSerLeuGluGluProSerLysAspTyrArg 80
             ||| ||||||||||||||||||| |||||||||||||||||||||||
Db   253 GGTNNNTGCTTAAGTGAAGGCTCTGATTNNNGTTTAGAAGAACCNAGCAAGGACTATAGA 312

Qy   81 GluAsnHisAsnAsnAsnLysLysAspHisSerAspSerAspAspGlnMetLeuValAla 100
             |||||| ||||||||||||||| |||||||||||||||||||||||

```
Db     313 GAGAATNNNNNTAATAATAAAAAAGATNNNAGTGACTCTGATGACCAAATGTTNNNNNCA 372

Qy     101 LysArgArgProSerSerAsnLeuAsnAsnAsnValArgGlyLysArgProLeuTrpHis 120
           |||||||||     |||||||||||||||||||||||||||||         |||||||||
Db     373 AAGCGCAGGNNATCATCAAACTTAAATAATAATGTTCGANNGAANNNNCCTCTATGGCAN 432

Qy     121 GluSerAspPheAlaValAspAsnValGlyAsnArgThrLeuArgArgArgArgLysVal 140
               |||||||||    |||    ||||||
Db     433 GNNNNTGATTTTGCTGNNGACNNTGTTGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 492

Qy     141 LysArgMetAlaValAspLeuProGlnAspIleSerAsnLysArgThrMetThrGlnPro 160

Db     493 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 552

Qy     161 ProGluGlyCysArgAspGlnAspMetAspSerAspArgAlaTyrGlnTyrGlnGluPhe 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     553 NNNNNNNNNTTGTAGAGATCAGGACATGGACAGTGATAGAGCCTACCAGTATCAAGAATTT 612

Qy     181 ThrLysAsnLysValLysLysArgLysLeuLysIleIleArgGlnGlyProLysIleGln 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     613 ACCAAGAACAAAGTCAAAAAAAGAAAGTTGAAAATAATCAGACAAGGACCAAAAATCCAA 672

Qy     201 AspGluGlyValValLeuGluSerGluGluThrAsnGlnThrAsnLysAspLysMetGlu 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     673 GATGAAGGAGTAGTTTTAGAAAGTGAGGAAACGAACCAGACCAATAAGGACAAAATGGAA 732

Qy     221 CysGluGluGlnLysValSerAspGluLeuMetSerGluSerAspSerSerSerLeuSer 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     733 TGTGAAGAGCAAAAAGTCTCAGATGAGCTCATGAGTGAAAGTGATTCCAGCAGTCTCAGC 792

Qy     241 SerThrAspAlaGlyLeuPheThrAsnAspGluGlyArgGlnGlyAspAspGluGlnSer 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     793 AGCACTGATGCTGGATTGTTTACCAATGATGAGGGAAGACAAGGTGATGATGAACAGAGT 852

Qy     261 AspTrpPheTyrGluLysGluSerGlyGlyAlaCysGlyIleThrGlyValValProTrp 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     853 GACTGGTTCTACGAAAAGGAATCAGGTGGAGCATGTGGTATCACTGGAGTTGTGCCCTGG 912

Qy     281 TrpGluLysGluAspProThrGluLeuAspLysAsnValProAspProValPheGluSer 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     913 TGGGAAAAGGAAGATCCTACTGAGCTAGACAAAAATGTACCAGATCCTGTCTTTGAAAGT 972

Qy     301 IleLeuThrGlySerPheProLeuMetSerHisProSerArgArgGlyPheGlnAlaArg 320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     973 ATCTTAACTGGTTCTTTTCCCCTTATGTCACACCCAAGCAGAAGAGGTTTCCAAGCTAGA 1032

Qy     321 LeuSerArgLeuHisGlyMetSerSerLysAsnIleLysLysSerGlyGlyThrProThr 340
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1033 CTCAGTCGCCTTCATGGAATGTCTTCAAAGAATATTAAAAAATCTGGAGGGACTCCAACT 1092

Qy     341 SerMetValProIleProGlyProValGlyAsnLysArgMetValHisPheSerProAsp 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1093 TCAATGGTACCCATTCCTGGCCCAGTGGGTAACAAGAGAATGGTTCATTTTTCCCCGGAT 1152

Qy     361 SerHisHisHisAspHisTrpPheSerProGlyAlaArgThrGluHisAspGlnHisGln 380
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1153 TCTCATCACCATGACCATTGGTTTAGCCCTGGGGCTAGGACAGAGCATGACCAGCATCAG 1212
```

```
Qy    381 LeuLeuArgAspAsnArgAlaGluArgGlyHisLysLysAsnCysSerValArgThrAla 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1213 CTNCTGAGAGATAATCGAGCTGAAAGAGGACACAAGAAAAATTGTTNNNNNNNNNNNNNNN 1272

Qy    401 SerArgGlnThrSerMetHisLeuGlySerLeuCysThrGlyAspIleLysArgArgArg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1273 NNNNNGCAAACAAGCATGCATTTAGGATCCTTATGCACGGGAGATATCAAACGGAGAAGA 1332

Qy    421 LysAlaAlaProLeuProGlyProThrThrAlaGlyPheValGlyGluAsnAlaGlnPro 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1333 AAAGCTGCACCTTTGCCTGGACCTACTACTGCAGGATTTGTAGGTGAAAATGCCCAGCCA 1392

Qy    441 IleLeuGluAsnAsnIleGlyAsnArgMetLeuGlnAsnMetGlyTrpThrProGlySer 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1393 ATCCTAGAAAATAATATTGGAAACCGAATGCTTCAGAATATGGGCTGGACGCCTGGGTCA 1452

Qy    461 GlyLeuGlyArgAspGlyLysGlyIleSerGluProIleGlnAlaMetGlnArgProLys 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1453 GGCCTTGGACGAGATGGCAAGGGGATCTCTGAGCCAATTCAAGCCATGCAGAGGCCAAAG 1512

Qy    481 GlyLeuGlyLeuGlyPheProLeuProLysSerThrSerAlaThrThrThrProAsnAla 500
          |||||||||||||||||||||||||||||||||||||||||||     ||||||||||||
Db   1513 GGATTAGGACTTGGATTTCCTCTACCAAAAGTACTTCCGCAACTGCTACCCCCAATGCA 1572

Qy    501 GlyLysSerAla 504
          ||||||||||||
Db   1573 GGAAAATCCGCC 1584
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 10:07:25 ; Search time 230 Seconds
                                             (without alignments)
                                             1546.027 Million cell updates/sec

Title: US-09-771-312-2
Perfect score: 2694
Sequence: 1 MEELVHDLVSALEESSEQAR..........GFPLPKSTSATTTPNAGKSA 504

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   UniProt_05.80:*
                 1: uniprot_sprot:*
                 2: uniprot_trembl:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                             SUMMARIES
                    %
Result   Query
   No.   Score Match Length DB   ID   Description
----
     1   2694   100.0   528   1   GPTC2_HUMAN   Q9nw75 homo sapien
     2   2694   100.0   528   2   Q5VYK7_HUMAN   Q5vyk7 homo sapien
     3   2310   85.7   527   1   GPTC2_MOUSE   Q7tqc7 mus musculu
     4 2009.5   74.6   504   2   Q5F3Y2_CHICK   Q5f3y2 gallus gall
     5   1813   67.3   376   2   Q5VYK8_HUMAN   Q5vyk8 homo sapien
     6 1587.5   58.9   414   2   Q4V7S5_XENLA   Q4v7s5 xenopus lae
     7   1538   57.1   375   2   Q9D3E7_MOUSE   Q9d3e7 mus musculu
     8 1513.5   56.2   410   2   Q6AY15_RAT   Q6ay15 rattus norv
     9 1283.5   47.6   561   2   Q4RRB2_TETNG   Q4rrb2 tetraodon n
    10   1008   37.4   216   2   Q6PIX0_HUMAN   Q6pix0 homo sapien
    11   939.5   34.9   408   2   Q5RJ37_BRARE   Q5rj37 brachydanio
    12   627.5   23.3   482   1   CN118_MOUSE   Q6pe65 mus musculu
    13   593.5   22.0   482   2   Q9H3M3_HUMAN   Q9h3m3 homo sapien
    14   545.5   20.2   453   1   CN118_HUMAN   Q9nwq4 homo sapien
    15   438   16.3   467   2   Q4RLV5_TETNG   Q4rlv5 tetraodon n
    16   430   16.0   107   2   Q9CSX3_MOUSE   Q9csx3 mus musculu
    17   320.5   11.9   221   2   Q9ULA8_HUMAN   Q9ula8 homo sapien
    18   197   7.3   928   2   Q6H4V9_ORYSA   Q6h4v9 oryza sativ
    19   177.5   6.6   1007   2   Q8VYR8_ARATH   Q8vyr8 arabidopsis
    20   177.5   6.6   1105   2   Q9M383_ARATH   Q9m383 arabidopsis
    21   170   6.3   812   2   Q6C233_YARLI   Q6c233 yarrowia li

```
22    169    6.3    742    2    Q6Z2C8_ORYSA         Q6z2c8 oryza sativ
23   167.5   6.2    767    1    YNW4_YEAST           P53866 saccharomyc
24   166.5   6.2   1469    2    Q5KKE0_CRYNE         Q5kke0 cryptococcu
25    165    6.1    732    2    Q5KCU3_CRYNE         Q5kcu3 cryptococcu
26    165    6.1    732    2    Q55IV9_CRYNE         Q55iv9 cryptococcu
27   164.5   6.1    781    2    Q9SF87_ARATH         Q9sf87 arabidopsis
28   163.5   6.1    346    2    Q5EB71_RAT           Q5eb71 rattus norv
29   162.5   6.0    505    2    Q59HE6_HUMAN         Q59he6 homo sapien
30   162.5   6.0    815    1    RBM5_HUMAN           P52756 homo sapien
31   162.5   6.0   1067    1    SFR14_MOUSE          Q8ch09 mus musculu
32    162    6.0    520    2    Q99KV9_MOUSE         Q99kv9 mus musculu
33    162    6.0    815    2    Q91YE7_MOUSE         Q91ye7 mus musculu
34   160.5   6.0    749    2    Q6DDU9_XENLA         Q6ddu9 xenopus lae
35   158.5   5.9   1469    2    Q55VU9_CRYNE         Q55vu9 cryptococcu
36   157.5   5.8    808    2    Q6BYP9_DEBHA         Q6byp9 debaryomyce
37   157.5   5.8    852    1    RBM10_RAT            P70501 rattus norv
38   157.5   5.8    853    2    Q8BTP8_MOUSE         Q8btp8 mus musculu
39   157.5   5.8    857    2    Q80U75_MOUSE         Q80u75 mus musculu
40   157.5   5.8    930    2    Q99KG3_MOUSE         Q99kg3 mus musculu
41    156    5.8    542    2    Q9NTB1_HUMAN         Q9ntb1 homo sapien
42    156    5.8    705    2    Q59UG4_CANAL         Q59ug4 candida alb
43    156    5.8    852    2    Q9BTX0_HUMAN         Q9btx0 homo sapien
44    156    5.8    853    2    Q5JRR2_HUMAN         Q5jrr2 homo sapien
45    156    5.8    929    1    RBM10_HUMAN          P98175 homo sapien
```

ALIGNMENTS

RESULT 1
GPTC2_HUMAN
ID   GPTC2_HUMAN   STANDARD;   PRT;   528 AA.
AC   Q9NW75; Q86YE7;
DT   05-JUL-2004 (Rel. 44, Created)
DT   05-JUL-2004 (Rel. 44, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   G patch domain containing protein 2.
GN   Name=GPATC2;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM 1).
RC   TISSUE=Embryo;
RX   PubMed=14702039; DOI=10.1038/ng1285;
RA   Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA   Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA   Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA   Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA   Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA   Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA   Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA   Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA   Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA   Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA   Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA   Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA   Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA   Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA   Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA   Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,

```
RA   Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA   Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA   Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA   Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA   Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA   Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA   Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA   Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA   Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA   Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT   "Complete sequencing and characterization of 21,243 full-length human
RT   cDNAs.";
RL   Nat. Genet. 36:40-45(2004).
RN   [2]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM 2).
RC   TISSUE=Lung, and Uterus;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9NW75-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9NW75-2; Sequence=VSP_010527, VSP_010528;
CC         Note=No experimental confirmation available;
CC   -!- SIMILARITY: Contains 1 G-patch domain.
CC   -----------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   -----------------------------------------------------------------------
DR   EMBL; AK001114; BAA91509.1; -; mRNA.
DR   EMBL; BC042193; AAH42193.1; -; mRNA.
DR   EMBL; BC063474; AAH63474.1; -; mRNA.
DR   Ensembl; ENSG00000092978; Homo sapiens.
DR   HGNC; HGNC:25499; GPATC2.
DR   InterPro; IPR000467; G_patch.
DR   Pfam; PF01585; G-patch; 1.
DR   SMART; SM00443; G_patch; 1.
DR   PROSITE; PS50174; G_PATCH; 1.
```

```
KW   Alternative splicing.
FT   DOMAIN        467    513      G-patch.
FT   VARSPLIC      367    376      VPIPGPVGNK -> ATNWTSEIPL (in isoform 2).
FT                                 /FTId=VSP_010527.
FT   VARSPLIC      377    528      Missing (in isoform 2).
FT                                 /FTId=VSP_010528.
FT   CONFLICT      220    220      G -> A (in Ref. 2; AAH63474).
FT   CONFLICT      225    225      D -> N (in Ref. 2; AAH42193).
SQ   SEQUENCE      528 AA;   58944 MW;   472143144700DC26 CRC64;
```

```
   Query Match            100.0%;  Score 2694;  DB 1;  Length 528;
   Best Local Similarity  100.0%;  Pred. No. 1.5e-153;
   Matches  504;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy         1 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        25 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 84

Qy        61 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        85 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 144

Qy       121 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       145 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 204

Qy       181 TKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       205 TKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 264

Qy       241 STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       265 STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 324

Qy       301 ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPD 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       325 ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPD 384

Qy       361 SHHHDHWFSPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRR 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       385 SHHHDHWFSPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRR 444

Qy       421 KAAPLPGPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPK 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       445 KAAPLPGPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPK 504

Qy       481 GLGLGFPLPKSTSATTTPNAGKSA 504
             ||||||||||||||||||||||||
Db       505 GLGLGFPLPKSTSATTTPNAGKSA 528
```

RESULT 2
Q5VYK7_HUMAN
```
ID   Q5VYK7_HUMAN PRELIMINARY;        PRT;    528 AA.
AC   Q5VYK7;
DT   01-FEB-2005 (TrEMBLrel. 29, Created)
DT   01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Novel protein.
GN   Name=RP11-361K17.1; ORFNames=RP11-361K17.1-001;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
```

```
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Griffiths C.;
RL   Submitted (MAY-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL354659; CAH70664.1; -; Genomic_DNA.
DR   EMBL; AC096641; CAH70664.1; JOINED; Genomic_DNA.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
SQ   SEQUENCE   528 AA;   58943 MW;   472143144700DC26 CRC64;
```

```
    Query Match              100.0%;  Score 2694;  DB 2;  Length 528;
    Best Local Similarity    100.0%;  Pred. No. 1.5e-153;
    Matches  504;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Qy      1 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 84

Qy     61 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 144

Qy    121 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 204

Qy    181 TKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205 TKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 264

Qy    241 STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    265 STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 324

Qy    301 ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPD 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    325 ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPD 384

Qy    361 SHHHDHWFSPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRR 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    385 SHHHDHWFSPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRR 444

Qy    421 KAAPLPGPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPK 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    445 KAAPLPGPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPK 504

Qy    481 GLGLGFPLPKSTSATTTPNAGKSA 504
          ||||||||||||||||||||||||
Db    505 GLGLGFPLPKSTSATTTPNAGKSA 528
```

```
RESULT 3
GPTC2_MOUSE
ID   GPTC2_MOUSE      STANDARD;       PRT;    527 AA.
AC   Q7TQC7; Q8BNJ9; Q8BPM1; Q8CDH9;
DT   05-JUL-2004 (Rel. 44, Created)
DT   05-JUL-2004 (Rel. 44, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   G patch domain containing protein 2.
GN   Name=Gpatc2;
OS   Mus musculus (Mouse).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORMS 2 AND 3).
RC   STRAIN=C57BL/6J; TISSUE=Eye, and Testis;
RX   MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM 1).
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
```

```
CC        Name=1;
CC          IsoId=Q7TQC7-1; Sequence=Displayed;
CC        Name=2;
CC          IsoId=Q7TQC7-2; Sequence=VSP_010529;
CC          Note=No experimental confirmation available;
CC        Name=3;
CC          IsoId=Q7TQC7-3; Sequence=VSP_010530;
CC          Note=No experimental confirmation available;
CC    -!- SIMILARITY: Contains 1 G-patch domain.
CC    --------------------------------------------------------------------------
CC    This Swiss-Prot entry is copyright. It is produced through a collaboration
CC    between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC    the European Bioinformatics Institute.  There are no  restrictions on  its
CC    use as long as its content is in no way modified and this statement is not
CC    removed.
CC    --------------------------------------------------------------------------
DR    EMBL; AK030026; BAC26744.1; -; mRNA.
DR    EMBL; AK053781; BAC35520.1; -; mRNA.
DR    EMBL; AK083471; BAC38928.1; -; mRNA.
DR    EMBL; BC054810; AAH54810.1; -; mRNA.
DR    Ensembl; ENSMUSG00000039210; Mus musculus.
DR    MGI; MGI:1915019; Gpatc2.
DR    InterPro; IPR000467; G_patch.
DR    Pfam; PF01585; G-patch; 1.
DR    SMART; SM00443; G_patch; 1.
DR    PROSITE; PS50174; G_PATCH; 1.
KW    Alternative splicing.
FT    DOMAIN      466    512      G-patch.
FT    VARSPLIC      1     23      Missing (in isoform 2).
FT                               /FTId=VSP_010529.
FT    VARSPLIC    388    425      DHWFSPGARTEHGQHQLLRDNRAERGHKKSCSLKTASR ->
FT                               E (in isoform 3).
FT                               /FTId=VSP_010530.
FT    CONFLICT    251    251      D -> Y (in Ref. 1; BAC26744).
FT    CONFLICT    367    367      S -> P (in Ref. 2).
SQ    SEQUENCE    527 AA;  58218 MW;  4F4F29FA56BE06B7 CRC64;
```

```
     Query Match                85.7%;  Score 2310;  DB 1;  Length 527;
     Best Local Similarity      84.9%;  Pred. No. 1.7e-130;
     Matches  428;  Conservative   35;  Mismatches   41;  Indels   0;  Gaps   0;

Qy        1 MEELVHDLVSALEESSEQARGGFAETGDHSRSISCPLKRQARKRRGRKRRSYNVHHPWET 60
            |||||||||||||||||||||||||||||||||:|||::|||||||||||||||||||||
Db       24 MEELVHDLVSALEESSEQARGGFAETGEHSRNLSCPLKRQARKRRGRKRRSYNVHHPWET 83

Qy       61 GHCLSEGSDSSLEEPSKDYRENHNNNKKDHSDSDDQMLVAKRRPSSNLNNNVRGKRPLWH 120
            |||||||||||||||||||||||| |:|||||| |||||||||||||||||:::||||| |||
Db       84 GHCLSEGSDSSLEEPSKDYREKHSNNKKDRSDSDDQMLVAKRRPSSNLSSSVRGKRLLWH 143

Qy      121 ESDFAVDNVGNRTLRRRRKVKRMAVDLPQDISNKRTMTQPPEGCRDQDMDSDRAYQYQEF 180
            |||||||::|||||||||||||||||||||||:|:|||| ||||||||||||| || ||
Db      144 ESDFAVDSLGNRTLRRRRKVKRMAVDLPQDVSSKRTMTQLPEGCRDQDMDNDRASQYPEF 203

Qy      181 TKNKVKKRKLKIIRQGPKIQDEGVVLESEETNQTNKDKMECEEQKVSDELMSESDSSSLS 240
            |:  ||||||||| || ||| |:|| |||||||  :|  |||:|| |||| |||||:||||
Db      204 TRKKVKKRKLKGIRPGPKTQEEGGVLESEERSQPNKDRMEYEEQKASDELRSESDTSSLS 263

Qy      241 STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGITGVVPWWEKEDPTELDKNVPDPVFES 300
            ||||||||||||||||||||||||||||||||||||:::| ||| |:||||||||
Db      264 STDAGLFTNDEGRQGDDEQSDWFYEKESGGACGIAGVVPWWEKDEPAELDTNLPDPVFES 323

Qy      301 ILTGSFPLMSHPSRRGFQARLSRLHGMSSKNIKKSGGTPTSMVPIPGPVGNKRMVHFSPD 360
            ||:||||||||| | ||||||||||| |||||||| | | ||: ||| ||||||||||
```

```
Db   324 ILSGSFPLMSHPGRGGFQARLSRLHGTPSKNIKKSSGAPPSMLSAPGPGSNKRMVHFSPD 383

Qy   361 SHHHDHWFSPGARTEHDQHQLLRDNRAERGHKKNCSVRTASRQTSMHLGSLCTGDIKRRR 420
         :| ||||||||||||| ||||||||||||||||:||::||||||||||||||||||||||
Db   384 AHRHDHWFSPGARTEHGQHQLLRDNRAERGHKKSCSLKTASRQTSMHLGSLCTGDIKRRR 443

Qy   421 KAAPLPGPTTAGFVGENAQPILENNIGNRMLQNMGWTPGSGLGRDGKGISEPIQAMQRPK 480
         |||||||||| || |||||||||||:|||||||:|||||||||||||:||:||:||:||||
Db   444 KAAPLPGPTAAGIVGENAQPILESNIGNRMLQSMGWTPGSGLGRDGRGIAEPVQAVQRPK 503

Qy   481 GLGLGFPLPKSTSATTTPNAGKSA 504
         |||||||||||:  :   | :|   |
Db   504 GLGLGFPLPKSSPTSPAPTSGNPA 527
```